



**HAL**  
open science

## The MUSE Hubble Ultra Deep Field Survey

Roland Bacon, Simon Conseil, David Mary, Jarle Brinchmann, Martin Shepherd, Mohammad Akhlaghi, Peter M. Weilbacher, Laure Piqueras, Lutz Wisotzki, David Lagattuta, et al.

► **To cite this version:**

Roland Bacon, Simon Conseil, David Mary, Jarle Brinchmann, Martin Shepherd, et al.. The MUSE Hubble Ultra Deep Field Survey. *Astronomy & Astrophysics - A&A*, 2017, 608, pp.A1. 10.1051/0004-6361/201730833 . hal-01678478

**HAL Id: hal-01678478**

**<https://hal.science/hal-01678478v1>**

Submitted on 6 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The MUSE *Hubble* Ultra Deep Field Survey

## I. Survey description, data reduction, and source detection<sup>★</sup>

Roland Bacon<sup>1</sup>, Simon Conseil<sup>1</sup>, David Mary<sup>2</sup>, Jarle Brinchmann<sup>3, 11</sup>, Martin Shepherd<sup>1</sup>, Mohammad Akhlaghi<sup>1</sup>, Peter M. Weilbacher<sup>4</sup>, Laure Piqueras<sup>1</sup>, Lutz Wisotzki<sup>4</sup>, David Lagattuta<sup>1</sup>, Benoit Epinat<sup>5, 6</sup>, Adrien Guerou<sup>5, 7</sup>, Hanae Inami<sup>1</sup>, Sebastiano Cantalupo<sup>8</sup>, Jean Baptiste Courbot<sup>1, 9</sup>, Thierry Contini<sup>5</sup>, Johan Richard<sup>1</sup>, Michael Maseda<sup>3</sup>, Rychard Bouwens<sup>3</sup>, Nicolas Bouché<sup>5</sup>, Wolfram Kollatschny<sup>10</sup>, Joop Schaye<sup>3</sup>, Raffaella Anna Marino<sup>8</sup>, Roser Pello<sup>5</sup>, Christian Herenz<sup>4</sup>, Bruno Guiderdoni<sup>1</sup>, and Marcella Carollo<sup>8</sup>

<sup>1</sup> Univ. Lyon, Univ. Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, 69230 Saint-Genis-Laval, France

e-mail: roland.bacon@univ-lyon1.fr

<sup>2</sup> Laboratoire Lagrange, CNRS, Université Côte d'Azur, Observatoire de la Côte d'Azur, CS 34229, 06304 Nice, France

<sup>3</sup> Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

<sup>4</sup> Leibniz-Institut für Astrophysik Potsdam, AIP, An der Sternwarte 16, 14482 Potsdam, Germany

<sup>5</sup> IRAP, Institut de Recherche en Astrophysique et Planétologie, CNRS, Université de Toulouse, 14 avenue Édouard Belin, 31400 Toulouse, France

<sup>6</sup> Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388 Marseille, France

<sup>7</sup> ESO, European Southern Observatory, Karl-Schwarzschild Str. 2, 85748 Garching bei Muenchen, Germany

<sup>8</sup> ETH Zurich, Institute of Astronomy, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland

<sup>9</sup> ICube, Université de Strasbourg – CNRS, 67412 Illkirch, France

<sup>10</sup> Institut für Astrophysik, Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

<sup>11</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal

Received 21 March 2017 / Accepted 25 July 2017

### ABSTRACT

We present the MUSE *Hubble* Ultra Deep Survey, a mosaic of nine MUSE fields covering 90% of the entire HUDF region with a 10-h deep exposure time, plus a deeper 31-h exposure in a single 1.15 arcmin<sup>2</sup> field. The improved observing strategy and advanced data reduction results in datacubes with sub-arcsecond spatial resolution (0".65 at 7000 Å) and accurate astrometry (0".07 rms). We compare the broadband photometric properties of the datacubes to HST photometry, finding a good agreement in zeropoint up to  $m_{AB} = 28$  but with an increasing scatter for faint objects. We have investigated the noise properties and developed an empirical way to account for the impact of the correlation introduced by the 3D drizzle interpolation. The achieved  $3\sigma$  emission line detection limit for a point source is 1.5 and  $3.1 \times 10^{-19}$  erg s<sup>-1</sup> cm<sup>-2</sup> for the single ultra-deep datacube and the mosaic, respectively. We extracted 6288 sources using an optimal extraction scheme that takes the published HST source locations as prior. In parallel, we performed a blind search of emission line galaxies using an original method based on advanced test statistics and filter matching. The blind search results in 1251 emission line galaxy candidates in the mosaic and 306 in the ultra-deep datacube, including 72 sources without HST counterparts ( $m_{AB} > 31$ ). In addition 88 sources missed in the HST catalog but with clear HST counterparts were identified. This data set is the deepest spectroscopic survey ever performed. In just over 100 h of integration time, it provides nearly an order of magnitude more spectroscopic redshifts compared to the data that has been accumulated on the UDF over the past decade. The depth and high quality of these datacubes enables new and detailed studies of the physical properties of the galaxy population and their environments over a large redshift range.

**Key words.** galaxies: distances and redshifts – galaxies: high-redshift – cosmology: observations – methods: data analysis – techniques: imaging spectroscopy – galaxies: formation

## 1. Introduction

In 2003 the *Hubble* Space Telescope (HST) performed a 1 Megasecond observation with its Advanced Camera for Surveys (ACS) in a tiny 11 arcmin<sup>2</sup> region located within the *Chandra* Deep Field South: the *Hubble* Ultra Deep Field (HUDF, Beckwith et al. 2006). The HUDF immediately became the deepest observation of the sky. This initial observation

was augmented a few years later with far ultraviolet images from ACS/SBC (Voyer et al. 2009) and with deep near ultraviolet (Teplitz et al. 2013) and near infrared imaging (Oesch et al. 2010; Bouwens et al. 2011; Ellis et al. 2013; Koekemoer et al. 2013) using the Wide Field Camera 3 (WFC3). These datasets have been assembled into the eXtreme Deep Field (XDF) by Illingworth et al. (2013). With an achieved sensitivity ranging from 29.1 to 30.3 AB mag, this emblematic field is still, fourteen years after the start of the observations, the deepest ever high-resolution image of the sky. Thanks to a large range of ancillary data taken with other telescopes,

<sup>★</sup> Based on observations made with ESO telescopes at the La Silla Paranal Observatory under programs 094.A-0289(B), 095.A-0010(A), 096.A-0045(A) and 096.A-0045(B).

including for example *Chandra* (Xue et al. 2011; Luo et al. 2017), XMM (Comastri et al. 2011), ALMA (Walter et al. 2016; Dunlop et al. 2017), *Spitzer*/IRAC (Labbé et al. 2015), and the VLA (Kellermann et al. 2008; Rujopakarn et al. 2016), the field is also covered at all wavelengths from X-ray to radio.

Such a unique data set has been central to our knowledge of galaxy formation and evolution at intermediate and high redshifts. For example, Illingworth et al. (2013) have detected 14 140 sources at  $5\sigma$  in the field including 7121 galaxies in the deepest (XDF) region. Thanks to the exceptional panchromatic coverage of the *Hubble* images (11 filters from 0.3 to 1.6  $\mu\text{m}$ ) it has been possible to derive precise photometric redshifts for a large fraction of the detected sources. In particular, the latest photometric redshift catalog of Rafelski et al. (2015) provides 9927 photometric redshifts up to  $z = 8.4$ . This invaluable collection of galaxies has been the subject of many studies spanning a variety of topics, including: the luminosity function of high redshift galaxies (e.g., McLure et al. 2013; Finkelstein et al. 2015; Bouwens et al. 2015; Parsa et al. 2016), the evolution of star formation rate with redshift (e.g., Ellis et al. 2013; Madau & Dickinson 2014; Rafelski et al. 2016; Bouwens et al. 2016; Dunlop et al. 2017), measurements of stellar mass (e.g., González et al. 2011; Grazian et al. 2015; Song et al. 2016), galaxy sizes (e.g., Oesch et al. 2010; Ono et al. 2013; van der Wel et al. 2014; Shibuya et al. 2015; Curtis-Lake et al. 2016) and dust and molecular gas content (e.g., Aravena et al. 2016b,a; Decarli et al. 2016b,a), along with probes of galaxy formation and evolution along the *Hubble* sequence (e.g., Conselice et al. 2011; Szomoru et al. 2011).

Since the release of the HUDF, a significant effort has been made with 8-m class ground-based telescopes to perform follow-up spectroscopy of the sources detected in the deep HUDF images. Rafelski et al. (2015) compiled a list of 144 high confidence ground-based spectroscopic redshifts from various instruments and surveys (see their Table 3): VIMOS-VVDS (Le Fèvre et al. 2004), FORS1&2 (Szokoly et al. 2004; Mignoli et al. 2005) VIMOS-GOODS (Vanzella et al. 2005–2009; Popesso et al. 2009; Balestra et al. 2010) and VIMOS-GMASS (Kurk et al. 2013). In addition, HST Grism spectroscopy provided 34 high-confidence spectroscopic redshifts: GRAPES (Daddi et al. 2005) and 3DHST (Morris et al. 2015; Momcheva et al. 2016). This large and long lasting investment in telescope time has thus provided 178 high-confidence redshifts in the HUDF area since 2004. Although the number of spectroscopic redshifts makes up only a tiny fraction (2%) of the 9927 photometric redshifts (hereafter photo- $z$ ), they are essential for calibrating photo- $z$  accuracy. In particular, by using the reference spectroscopic sample, Rafelski et al. (2015) found that their photo- $z$  measurements achieved a low scatter (less than 0.03 rms in  $\sigma_{\text{NMAD}}$ ) with a reduced outlier fraction (2.4–3.8%).

However, this spectroscopic sample is restricted to bright objects (the median  $F775W$  AB magnitude of the sample is 23.7, with only 12% having  $AB > 25$ ) at low redshift: the sample distribution peaks at  $z \approx 1$  and only a few galaxies have  $z > 2$ . The behavior of spectrophotometric methods at high  $z$  and faint magnitude is therefore poorly known. Given that most of the HUDF galaxies fall in this regime (96% of the Rafelski et al. 2015, sample has  $AB > 25$  and 55% has  $z > 2$ ), it would be highly desirable to obtain a larger number of high-quality spectra in this magnitude and redshift range.

Besides calibrating the photo- $z$  sample, though, there are other important reasons to increase the number of sources in the UDF with high quality spectroscopic information. Some key astrophysical properties of galaxies can only be measured from

spectroscopic information, including kinematics of gas and stars, metallicity, and the physical state of gas. Environmental studies also require a higher redshift accuracy than those provided by photo- $z$  estimates.

The fact that only a small fraction of objects seen in the HST images (representing the tip of the iceberg of the galaxy population) have spectroscopic information shows how difficult these measurements are. In particular, the current state-of-the-art multi-object spectrographs perform well when observing the bright end of galaxy population over wide fields. But, despite their large multiplex, they are not well adapted to perform deep spectroscopy in very dense environments. An exhaustive study of the UDF galaxy population with these instruments would be prohibitively expensive in telescope time and very inefficient. Thus, by practical considerations, multi-object spectroscopy is restricted to studying preselected samples of galaxies. Since pre-selection implies that only objects found in broadband deep imaging will be selected, this technique leaves out potential emission-line only galaxies with faint continua.

Thankfully, with the advent of MUSE, the Multi Unit Spectroscopic Explorer at the VLT (Bacon et al. 2010) the state of the art is changing. As expressed in the original MUSE science case (Bacon et al. 2004), one of the project's major goals is to push beyond the limits of the present generation of multi-object spectrographs, using the power of integral field spectroscopy to perform deep spectroscopic observations in *Hubble* deep fields.

During the last MUSE commissioning run (Bacon et al. 2014) we performed a deep 27-h integration in a 1 arcmin<sup>2</sup> region located in the *Hubble* Deep Field South (hereafter HDFS) to validate MUSE's capability in performing a blind spectroscopic survey. With this data we were able to improve the number of known spectroscopic redshifts in this tiny region by an order of magnitude (Bacon et al. 2015). This first experiment not only effectively demonstrated the unique capabilities of MUSE in this context, but has also led to new scientific results: the discovery of extended Ly $\alpha$  halos in the circumgalactic medium around high redshift galaxies (Wisotzki et al. 2016), the study of gas kinematics (Contini et al. 2016), the investigation of the faint-end of the Ly $\alpha$  luminosity function (Drake et al. 2017a), the measurement of metallicity gradients (Carton et al. 2017) and the properties of galactic winds at high  $z$  (Finley et al. 2017a).

The HDFS observations also revealed 26 Ly $\alpha$  emitting galaxies that were not detected in the HST WFPC2 deep broadband images, demonstrating that continuum-selected samples of galaxies, even at the depth of the *Hubble* deep fields, do not capture the complete galaxy population. This collection of high equivalent width Ly $\alpha$  emitters found in the HDFS indicates that such galaxies may be an important part of the low-mass, high-redshift galaxy population. However, this first investigation in the HDFS was limited to a small 1 arcmin<sup>2</sup> field of view and will need to be extended to other deep fields before we can assess its full importance.

After the HDFS investigation, the next step was to start a more ambitious program on the *Hubble* Ultra Deep Field. This project was conducted as one of the guarantee time observing (GTO) programs given by ESO in return for the financial investment and staff effort brought by the Consortium to study and build MUSE. This program is part of a wedding cake approach, consisting of the shallower MUSE-Wide survey in the CDFS and COSMOS fields (Herenz et al. 2017) covering a wide area, along with a deep and ultra-deep survey in the HUDF field covering a smaller field of view.

This paper (hereafter Paper I) is the first paper of a series that describes our investigation of the HUDF and assesses the

science results. Paper I focuses on the details of the observations, data reduction, performance assessment and source detection. In Paper II (Inami et al. 2017) we describe the redshift analysis and provide the source catalog. In Paper III (Brinchmann et al. 2017) we investigate the photometric redshifts properties of the sample. The properties of CIII] emitters as Ly $\alpha$  alternative for redshift confirmation of high- $z$  galaxies are discussed in Paper IV (Maseda et al. 2017). In Paper V (Gu erou et al. 2017) we obtain spatially resolved stellar kinematics of galaxies at  $z \approx 0.2$ – $0.8$  and compare their kinematical properties with those inferred from gas kinematics. The faint end of the Ly $\alpha$  luminosity function and its implication for reionization are presented in Paper VI (Drake et al. 2017b). The properties of Fe ii\* emission, as tracer of galactic winds in star-forming galaxies is presented in Paper VII (Finley et al. 2017b). Extended Ly $\alpha$  haloes around individual Ly $\alpha$  emitters are discussed in Paper VIII (Leclercq et al. 2017). The first measurement of the evolution of galaxy merger fraction up to  $z \approx 6$  is presented in Paper IX (Ventou et al. 2017) and a detailed study of Ly $\alpha$  equivalent widths properties of the Ly $\alpha$  emitters is discussed in Paper X (Hashimoto et al. 2017).

The paper is organized as follows. After the description of the observations (Sect. 2), we explain the data reduction process in detail (Sect. 3). The astrometry and broadband photometric performances are discussed in Sect. 4. We then present the achieved spatial and spectral resolution (Sect. 5), including an original method to derive the spatial PSF when there is no point source in the field. Following that, we investigate in Sect. 6 the noise properties in detail and derive an estimate of the limiting emission line source detection. Finally, we explain how we perform source detection and describe an original blind search algorithm for emission line objects (Sect. 7). A summary concludes the paper.

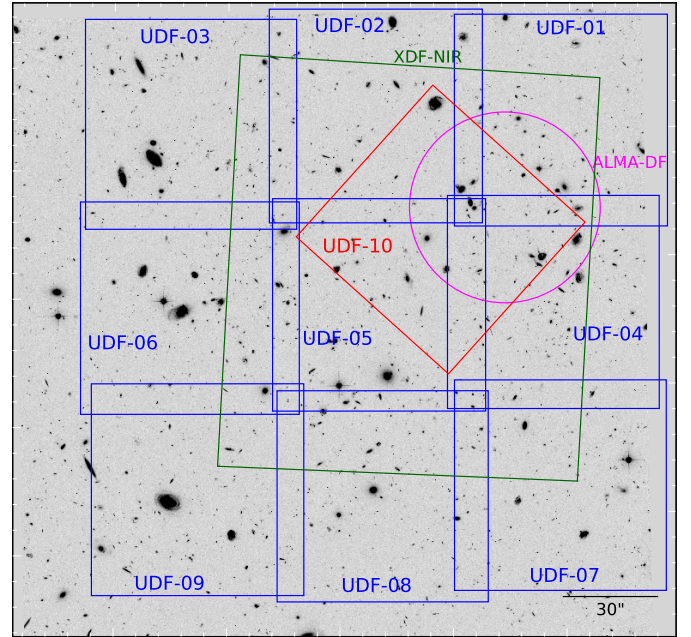
## 2. Observations

The HUDF was observed over eight GTO runs over two years: September, October, November and December 2014, August, September, October, and December 2015 and February 2016. A total of 137 h of telescope in dark time and good seeing conditions have been used for this project. This is the equivalent to 116 h of open shutter time which translates to 85% efficiency when including the overheads.

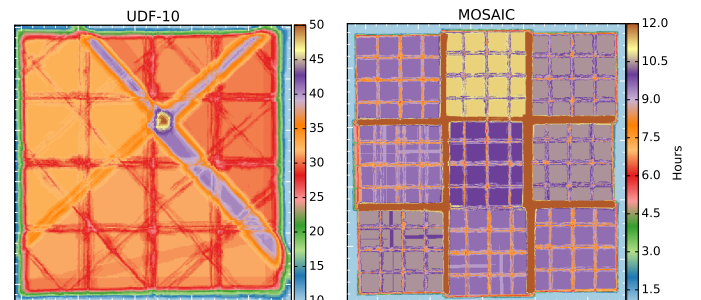
### 2.1. The medium deep mosaic field

We covered the HUDF region with a mosaic of nine MUSE fields (UDF-01 through UDF-09, respectively) oriented at a PA of  $-42^\circ$  as shown in Fig. 1. Each MUSE field is approximately a square  $1 \times 1$  arcmin<sup>2</sup> in area. The dithering pattern used is similar to the HDFS observation scheme (Bacon et al. 2015): that is, a set of successive  $90^\circ$  instrument rotations plus random offsets within a  $2''$  square box.

Given its declination ( $-27^\circ 47' 29''$ ), the UDF transits very close to zenith in Paranal. When approaching zenith, the rotation speed of the instrument optical derotator increases significantly and its imperfect centering produces a non negligible wobble. However, MUSE has the ability to perform secondary guiding, using stars positioned in a circular ring around the field of view. Image of these stars are affected by the derotator wobble in the same way as the science field, so their shapes can be used to correct for the extra motion. The use of a good slow-guiding star is therefore very important in maintaining field-centering during an exposure, in order to get the best spatial resolution. Thus, the location of each field in the mosaic was optimized to not only



**Fig. 1.** Field location and orientation for the mosaic (UDF01–09, in blue) and UDF10 (in red) fields, overlaid on the HST ACS  $F775W$  image. The green rectangle indicates the XDF/HUDF09/HUDF12 region containing the deepest near-IR observations from the HST WFC3/IR camera. The magenta circle displays the deep ALMA field from the ASPECS pilot program (Walter et al. 2016). North is located  $42^\circ$  clockwise from the vertical axis.



**Fig. 2.** Final exposure map images (averaged over the full wavelength range) in hours for the udf-10 and mosaic fields. The visible stripes correspond to regions of lower integration due to the masking process (see Sect. 3.1.3).

provide a small overlap with adjacent fields but also to keep the selected slow-guiding star within the slow-guiding region during the rotation+dither process. Unfortunately, only a fraction of the fields have an appropriate slow-guiding star within their boundaries (UDF-02, 04, 07, and 08). Therefore, we preferentially observed these fields when the telescope was near zenith, while the others were observed when the zenith angle was larger than  $10^\circ$ .

The integration time for each exposure was 25 min. This is long enough to reach the sky-noise-limited regime, even in the blue range of the spectrum, but still short enough to limit the impact of cosmic rays. Including the overheads it is possible to combine two exposures into an observing block spanning approximately 1 h. A total of 227 25-min exposures were performed in good seeing conditions. A few exposures were repeated when the requested conditions were not met (e.g., poor seeing or cirrus absorption). As shown in Fig. 2 and taking

into account a few more exposures that were discarded for various reasons during the data reduction process (see Sect. 3), the mosaic field achieves a depth of  $\approx 10$  h over a contiguous area of  $9.92 \text{ arcmin}^2$  within a rectangle approximately  $3.15' \times 3.15'$  in shape.

## 2.2. The udf-10 ultra deep field

In addition to the mosaic, we also performed deeper observations of a single  $1' \times 1'$  field, called UDF-10. The field location<sup>1</sup> was selected to be in the deepest part of the XDF/HUDF09/HUDF12 area and to overlap as much as possible with the deep ALMA pointing from the ASPECS pilot program (Walter et al. 2016). A different PA of  $0^\circ$  was deliberately chosen to better control the systematics. Specifically, when this field is combined with the overlapping mosaic fields (at a PA of  $-42^\circ$ ), the instrumental slice and channel orientation with respect to the sky is different. This helps to break the symmetry and minimize the small systematics that are left by the data reduction process. Care was taken to have a bright star within the slow-guiding field in order to obtain the best possible spatial resolution, even when the field transits near zenith. Because of this additional constraint, the field only partially overlaps with the deep ALMA pointing. The resulting location is shown in Fig. 1.

Given that GTO observations are conducted in visitor mode and not in service mode, we performed an equivalent GTO queue scheduling within all GTO observing programs. A fraction of the best seeing conditions were used for this field. During observation, we used the same dithering strategy and individual exposure time as for the mosaic, obtaining a total of 51 25-min exposures.

In the following we call udf-10 the combination of UDF-10 with the overlapping mosaic fields (UDF-01, 02, 04, and 05). udf-10 covers an area of  $1.15 \text{ arcmin}^2$  and reaches a depth of 31 h (Fig. 2). Such a depth is comparable to the 27 h reached by the HDFs observations (Bacon et al. 2015). However, as we will see later, the overall quality is much better thanks to the best observing conditions, an improved observational strategy and refined data reduction process.

## 3. Data reduction

Performing reductions on such a large data set (278 science exposures) is not a negligible task, but the control and minimization of systematics is extremely important since we want to make optimal use of the depth of the data. The overall process for the UDF follows the data reduction strategy developed for the HDFs (Bacon et al. 2015) but with improved processes and additional procedures (see Conseil et al. 2017). It consists of two major steps: the production of a datacube from each individual exposure and the combination of the datacubes to produce the final mosaic and udf-10 datacubes. These steps are described in the following sections.

### 3.1. Data reduction of individual exposures

#### 3.1.1. From the raw science data to the first pixtable

We first run the raw science data through the MUSE standard pipeline version 1.7 dev (Weilbacher et al., in prep.). The individual exposures are processed by the *scibasic* recipe which used the corresponding daily calibrations (flatfields, bias, arc lamps,

<sup>1</sup> The udf-10 field center is at  $\alpha_{J2000} = 03^{\text{h}}32^{\text{m}}38.7^{\text{s}}$ ,  $\delta_{J2000} = -27^\circ46'44''$ .

twilight exposures) and geometry table (one per observing run) to produce a table (hereafter called *pixtable*) containing all pixel information: location, wavelength, photon count and an estimate of the variance. Bad pixels corresponding to known CCD defects (columns or pixels) are also masked at this time. For each exposure we use the *illumination* exposure to correct for flux variations at the slices edges due to small temperature changes between the morning calibration exposures and the science exposures. From the adjacent *illumination* exposures taken before and after the science, we select the one nearest in temperature.

The pipeline recipe *scipost* is then used to perform astrometric and flux calibrations on the *pixtable*. We use a single reference flux calibration response for all exposures, created in the following way. All flux calibration responses, obtained over all nights, are scaled to the same mean level to remove transparency variations. Then, we take the median of the stack to produce the final reference response. We note that no sky subtraction is performed at this stage because we use the sky flux to perform self-calibration on each exposure.

A datacube is then created with the *makecube* pipeline recipe, using the default 3D drizzling interpolation process. Each exposure needs to be precisely recentered to correct for the derotator wobble. Unlike the HDFs observations, only a few UDF fields have bright point sources that can be used to compute this offset. We have therefore developed an original method to derive precise offset values with respect to the HST reference images. This is described in detail in Sect. 5.1. The computed ( $\Delta\alpha, \Delta\delta$ ) offset values are then applied to the *pixtable*, which is then ready for the self-calibration process.

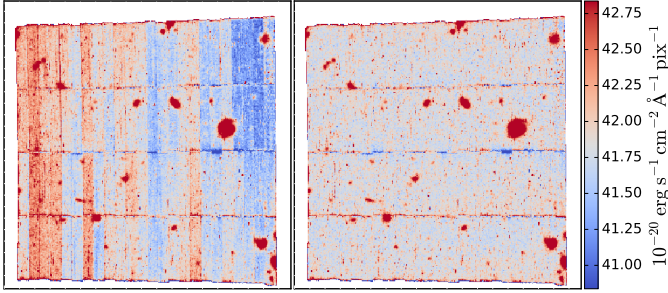
#### 3.1.2. Self calibration

Although the standard pipeline is efficient at removing most of the instrumental signatures, one can still see a low-level footprint of the instrumental slices and channels. This arises from a mix of detector instabilities and imperfect flatfielding, which are difficult to correct for with standard calibration exposures. We therefore use a self-calibration procedure<sup>2</sup>, similar in spirit to the one used for the HDFs (Bacon et al. 2015) but enhanced to produce a better correction. It is also similar to the CubeFIX flat-fielding correction method, part of the CubeExtractor package developed by Cantalupo (in prep.) and used, for instance, in Borisova et al. (2016, see therein for a short description) but it works directly on the *pixtable*. Compared to the HDFs version, the major changes in the new procedure are to perform polychromatic correction and to use a more efficient method to reject outliers.

The procedure starts by masking all bright objects in the data. The mask we use is the same for all exposures, calculated from the white light image of the rough, first-pass datacube of the combined UDF data set. The method works on 20 wavelength bins of 200–300 Å. These bins have been chosen so that their edges do not fall on a sky line. The median flux of each slice<sup>3</sup> is computed over the wavelength range of the bin, using only

<sup>2</sup> The self-calibration procedure is part of the MPDAF software (Piqueras et al. 2017): the MUSE Python Data Analysis Framework. It is an open-source (BSD licensed) Python package, developed and maintained by CRAL and partially funded by the ERC advanced grant 339659-MUSICOS. It is available at <https://git-cral.univ-lyon1.fr/MUSE/mpdaf>

<sup>3</sup> The slices are the thin mirrors of the MUSE image slicer which perform the reformatting of the entrance field of view into a pseudo slit located at the spectrograph input focal plane.



**Fig. 3.** Self-calibration on individual exposures. The reconstructed white light image of a single exposure, highly stretched around the mean sky value, is shown before (*left panel*) and after (*right panel*) the self calibration process.

the unmasked voxels<sup>4</sup> in the slice. Individual slices flux are then offset to the mean flux of all slices and channels over the same wavelength bin. Outliers are rejected using  $15\sigma$  clipping based on a the median absolute deviation (MAD). As shown in Fig. 3, the new self calibration is very efficient in removing the remaining flatfielding defects and other calibration systematics.

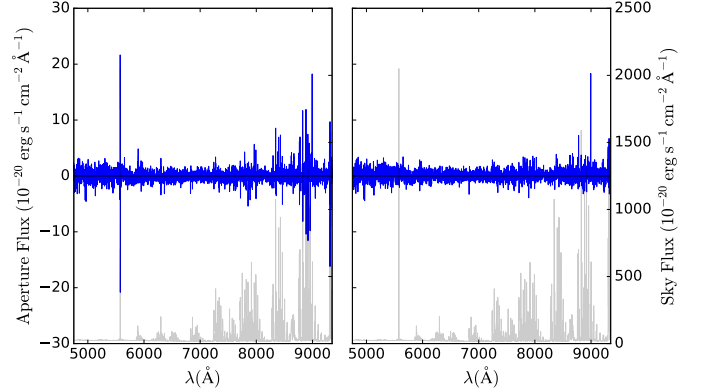
### 3.1.3. Masking

Some dark or bright regions at the edges of each slice stack (hereafter called inter-stack defects) can be seen as thin, horizontal strips in Fig. 3. These defects are not corrected by standard flat-fielding or through self-calibration and appear only in deep exposures of the empty field. It is important to mask them because otherwise the combinations of many exposures at different instrumental rotation angles and with various on-sky offsets will impact a broad region on the final datacubes.

To derive the optimum mask, we median-combine all exposures, irrespective of the field, projected on an instrumental grid (i.e., we stack based on fixed pixel coordinates instead of the sky’s world coordinate system). In such a representation, the instrumental defects are always at the same place, while sky objects move from place to place according to the dithering process. The resulting mask identifies the precise locations of the various defects on the instrumental grid. This is used to build a specific *bad pixel table* which is then added as input to the standard *scibasic* pipeline recipe.

In principle, to mask the inter-stack region one can simply produce a datacube using this additional *bad pixel table* with the *scibasic* and *scipost* recipes. However, the 3D drizzle algorithm used in *scipost* introduces additional interpolation effects which prevents perfect masking. To improve the inter-stack masking, we run the *scibasic* and *scipost* recipes twice: the first time without using the specific *bad pixel table*, and the second time with it. Using the output of the “bad-pixel” version of the cube, we derive a new, 3D mask which we apply to the original cube, effectively removing the inter-stack bad data.

Even after this masking, a few exposures had some unique problems which required additional specific masking. This was the case for 2 exposures impacted by Earth satellite trails, and for 9 exposures that show either high dark levels in channel 1 or important bias residuals in channel 6. An individual mask was built and applied for each of these exposures. The impact of all masking can be easily seen in Fig. 2 where the stripes with lower integration time show up in the exposure maps.



**Fig. 4.** Spectrum extracted from a  $1''$  diameter aperture in an empty region of a single exposure datacube, before (*left panel*) and after (*right panel*) the use of ZAP. The mean sky spectrum is shown in light gray.

### 3.1.4. Sky subtraction

The recentered and self-calibrated *pixtable* of each exposure is then sky subtracted, using the *scipost* pipeline recipe with sky subtraction enabled, and a new datacube is created on a fixed grid. For the mosaic field, we pre-define a single world coordinate system (with a PA of  $-42^\circ$ ) covering the full mosaic region, and each of the nine MUSE fields (UDF-1 through 9) is projected onto the grid. For the udf-10 a different grid is used (PA =  $0^\circ$ ). Based on the overlap region, fields UDF-1, 2, 4, 5 and 10 are projected onto this grid.

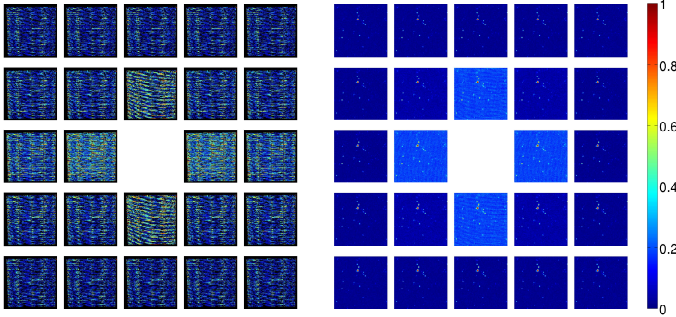
We then used ZAP (Soto et al. 2016), the principal component analysis enhanced sky subtraction software developed for MUSE datacubes. As shown in Fig. 4, ZAP is very efficient at removing the residuals left over by the standard pipeline sky subtraction recipe. The computed inter-stack 3D mask is then applied to the resulting datacube.

### 3.1.5. Variance estimation

Variance estimation is a critical step that is used to evaluate the achieved signal-to-noise ratio and to perform source extraction and detection, as we will see later in Sect. 7. The pipeline first records an estimate of the variance at each voxel location, using the measured photon counts as a proxy for the photon noise variance and adding the read-out detector variance. This variance estimate is then propagated accurately along each step of the reduction, taking into account the various linear transformations that are performed on the *pixtable*. However, even after accounting for these effects, there are still problems with the variance estimates.

The first problem is that the estimate is noisy, given that the random fluctuations around the unknown mean value are used in place of the mean itself for each pixel. The second problem is related to the interpolation used to build the regular grid of the datacube from the non-regular *pixtable* voxels. This interpolation creates correlated noise in the output datacube as can be seen in Fig. 5. To take into account this correlation, one should in principle propagate both the variance information and the covariance matrix, instead of just the variance as the pipeline does. However, this covariance matrix is far too large ( $\approx 125$  times the datacube size, even if we limit it to pixels within the seeing envelope and 5 pixels along the spectral axis) and thus cannot be used in practice.

<sup>4</sup> Voxel: volume sampling element ( $0''.2 \times 0''.2 \times 1.25 \text{ \AA}$ ).



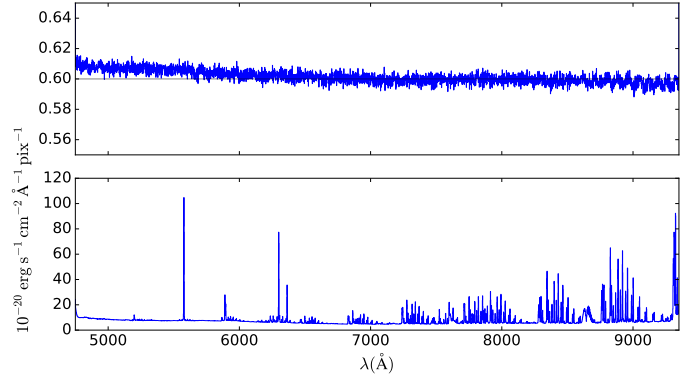
**Fig. 5.** Spatially correlated properties in the MUSE udf-10 datacube after drizzle interpolation. Each image shows the correlation between spectra and their  $\pm 1$ ,  $\pm 2$  spatial neighbors. The correlation image is shown for a single exposure datacube (*left panel*) and for the combined datacube (*right panel*). Note that the correlation was performed on the blue part of the spectrum to avoid the OH lines region.

The consequence is that the pipeline-propagated variance for a single exposure exhibits strong oscillations along both the spatial and spectral axes. When combining multiple datacube exposures into one, the spatial and spectral structures of the variance are reasonably flat, since the various oscillations cancel out in the combination. However, because we ignore the additional terms of the covariance matrix, the pipeline-propagated noise estimation is still wrong in terms of its absolute value. Ideally, we should then work only with *pixtable* to avoid this effect. However, this is difficult in practice because most of signal processing and visualization routines (e.g., Fast Fourier Transform) require a regularly sampled array.

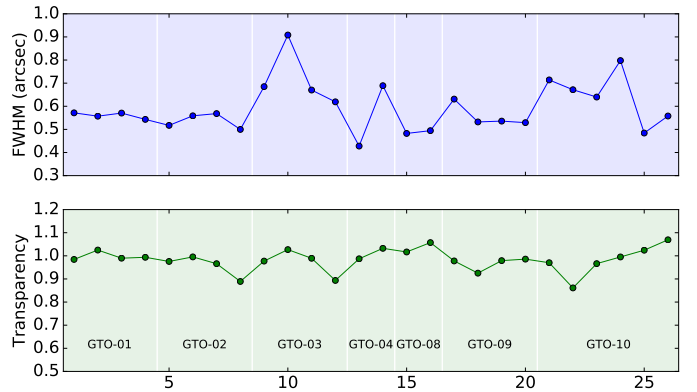
To face this complex problem<sup>5</sup> we have adopted a scheme to obtain a more realistic variance estimate for faint objects where the dominant source of noise is the sky. In this case the variance is a function of wavelength only. For faint objects, we will always sum up the flux over a number of spatial and spectral bins, such as (for example) a  $1''$  diameter aperture to account for atmospheric seeing and a few  $\text{\AA}$  along the spectral axis. As can be seen in Fig. 5, the correlation impact is strongly driven by a pixel's immediate neighbors but decreases very rapidly at larger distances. The same behavior is found along the spectral axis. Thus, if the 3D aperture size is large enough with respect to the correlation size, the variance of the aperture-summed signal should be equal to the original variance prior to resampling.

As a test to reconstruct the original pre-resampling variances, we perform the following experiment. We start with a *pixtable* that produces an individual datacube, which will later be combined with the other exposures. We fill this *pixtable* with perfect Gaussian noise (with a mean of zero and a variance of 1) and then produce a datacube using the standard pipeline 3D drizzle. As expected, the pixel-to-pixel variance of this test datacube is less than 1 because of the correlation. The actual value depends on the *pixfrac* drizzle parameter related to the number of neighboring voxels which are used in the interpolation process. With our *pixfrac* of 0.8, we measure a pixel-to-pixel standard deviation of 0.60 in our experimental datacube. This value is almost independent of wavelength as can be seen in Fig. 6. The ratio  $\frac{1}{0.60}$  is then the correction factor that needs to be applied to the pixel-to-pixel standard deviation.

To overcome the previously mentioned problem of noise in the pipeline-propagated variance estimator, we re-estimate the



**Fig. 6.** Example of estimated standard deviation corrected for correlation effects (see text) in one exposure. *Top*: pixel-to-pixel standard deviation of the experimental noisy datacube and adopted correction factor. *Bottom*: pixel-to-pixel standard deviation of a real one-exposure datacube after correcting for correlation effects.



**Fig. 7.** Computed variation of FSF FWHM at  $7750 \text{ \AA}$  (*top panel*) and transparency (*bottom panel*) for all exposures of the UDF-04 field obtained in seven GTO runs.

pixel-to-pixel variance directly from each datacube. We first mask the bright sources and then measure the median absolute deviation for each wavelength. The resulting standard deviation is then multiplied by the correction factor to take into account the correlations. An example is shown in Fig. 6. Note however that this variance estimate is likely to be wrong for bright sources which are no longer dominated by the sky noise, and thus no longer have spatially constant variances. Given the focus of the science objectives, this is not considered a major problem in this work.

### 3.1.6. Exposure properties

In the final step before combining all datacubes, we evaluate some important exposure properties, such as their achieved spatial resolution and absolute photometry. We use the tool described in Sect. 5.1 to derive the FWHM of the Moffat PSF fit and the photometric correction of the MUSE exposure that gives the best match with the HST broadband images. An example of the evolution of the spatial resolution and photometric properties of the UDF-04 field is given in Fig. 7. The statistics of exposure properties for all fields is given in Table 1.

Control quality pages have been produced for all 278 individual exposures displaying various images, spectra and indicators for the steps of the data reduction. They were all visually

<sup>5</sup> Note that this variance behavior is not specific to these observations but is currently present in all MUSE datacubes provided by the pipeline.

**Table 1.** Observational properties of UDF fields.

Field	$N$	$F_m$	$F_\sigma$	$F_{\min}$	$F_{\max}$	$P_m$	$P_\sigma$	$P_{\min}$	$P_{\max}$	$F_b$	$F_r$
01	26	0.62	0.12	0.46	1.01	0.98	0.07	0.78	1.08	0.71	0.57
02	28	0.60	0.11	0.42	0.82	1.01	0.02	0.97	1.06	0.69	0.56
03	24	0.61	0.09	0.46	0.76	1.01	0.03	0.95	1.09	0.72	0.55
04	26	0.59	0.10	0.43	0.91	0.98	0.05	0.86	1.07	0.72	0.54
05	25	0.63	0.08	0.46	0.79	0.95	0.07	0.77	1.01	0.72	0.58
06	24	0.62	0.06	0.55	0.78	0.99	0.03	0.95	1.07	0.71	0.56
07	24	0.59	0.08	0.46	0.72	0.99	0.03	0.93	1.07	0.68	0.54
08	24	0.63	0.08	0.43	0.81	0.98	0.04	0.86	1.09	0.72	0.58
09	26	0.67	0.07	0.56	0.83	0.98	0.03	0.90	1.03	0.76	0.62
10	51	0.60	0.08	0.42	0.77	1.02	0.04	0.86	1.09	0.71	0.55

**Notes.** For each field the number of individual exposures ( $N$ ) is given, along with some statistics of the FWHM (arcsec) of the estimated point spread function (FSF) at 7750 Å: the mean ( $F_m$ ), standard deviation ( $F_\sigma$ ), and min ( $F_{\min}$ ) and max ( $F_{\max}$ ) values. Statistics of the relative photometric properties of each field are also given: the mean ( $P_m$ ), standard deviation ( $P_\sigma$ ), and min ( $P_{\min}$ ) and max ( $P_{\max}$ ) values. Additionally, the fit FWHM (in arcsec) of the combined datacube FSF is given at blue ( $F_b$  at 4750 Å) and red ( $F_r$  at 9350 Å) wavelengths.

inspected, and remedy actions were performed for the identified problems.

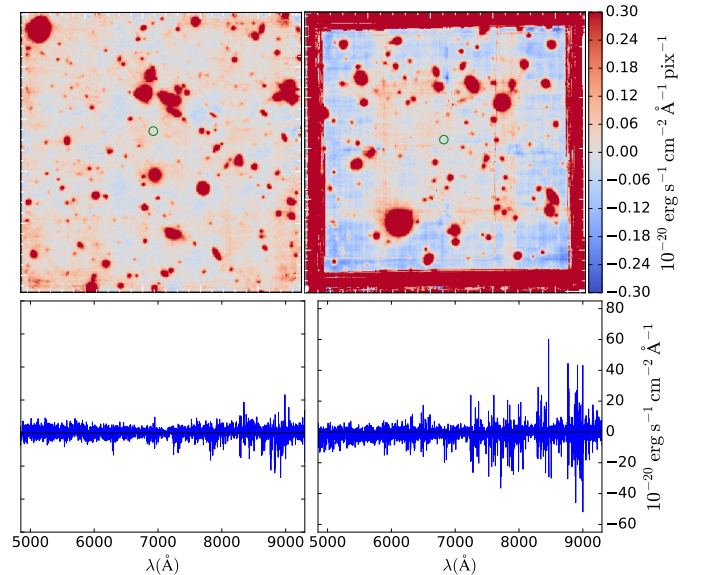
### 3.2. Production of the final datacubes

The 227 datacubes of the mosaic were combined, using the estimated flux corrections computed from a comparison with the reference HST image (see Sect. 5.1). We perform an average on all voxels, after applying a 5 sigma-clipping based on a robust median absolute deviation estimate to remove outliers. Except in the region of overlap between adjacent fields, or at the edges of the mosaic, each final voxel is created from the average of  $\approx 23$  voxels. The corrected variance is also propagated and an exposure map datacube is derived (see Fig. 2). The achieved median depth is 9.6 h. We also save the statistics of detected outliers to check if specific regions or exposures have been abnormally rejected. The resulting datacube is saved as a 25 GB multi-extension FITS file with two extensions: the data and the estimated variance. Each extension contains  $(n_x, n_y, n_\lambda) = 947 \times 945 \times 3681 = 3.29 \times 10^9$  voxels.

The same process is applied to the 51 UDF-10 proper datacubes plus the 105 overlapping mosaic datacubes (fields 01, 02, 04, and 05) projected onto the same grid. We note that four exposures with poor spatial resolution ( $FWHM > 0''.9$ ) have been removed from the combination. In this case,  $\approx 74$  voxels are averaged for each final voxel, leading to a median depth of 30.8 h (Fig. 2). The resulting 2.9 GB datacube contains  $(n_x, n_y, n_\lambda) = 322 \times 323 \times 3681 = 3.8 \times 10^8$  voxels. Note that the datacubes presented in this paper have the version 0.42.

To ensure that there is no background offset, we subtract the median of each monochromatic image from each cube, after proper masking of bright sources. The subtracted offsets are small:  $0.02 \pm 0.03 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ . The reconstructed white light images for the two fields, obtained simply by averaging over all wavelengths, are shown in Fig. 9.

To show the progress made since the HDFS publication (Bacon et al. 2015), we present in Fig. 8 a comparison between the HDFS cube and the udf-10 cube which achieves a similar depth. There are obvious differences: the bad-edge effect present in HDFS has now disappeared, the background is much flatter in the udf-10 field, while the HDFS shows negative and positive large scale fluctuations. The sky emission line residuals are also reduced as shown in the background spectra comparison. One



**Fig. 8.** Visual comparison between udf-10 (left) and HDFS (right) datacubes. White-light images are displayed in the top panels and examples of spectra extracted in an empty central region (green circle) are displayed in the bottom panels.

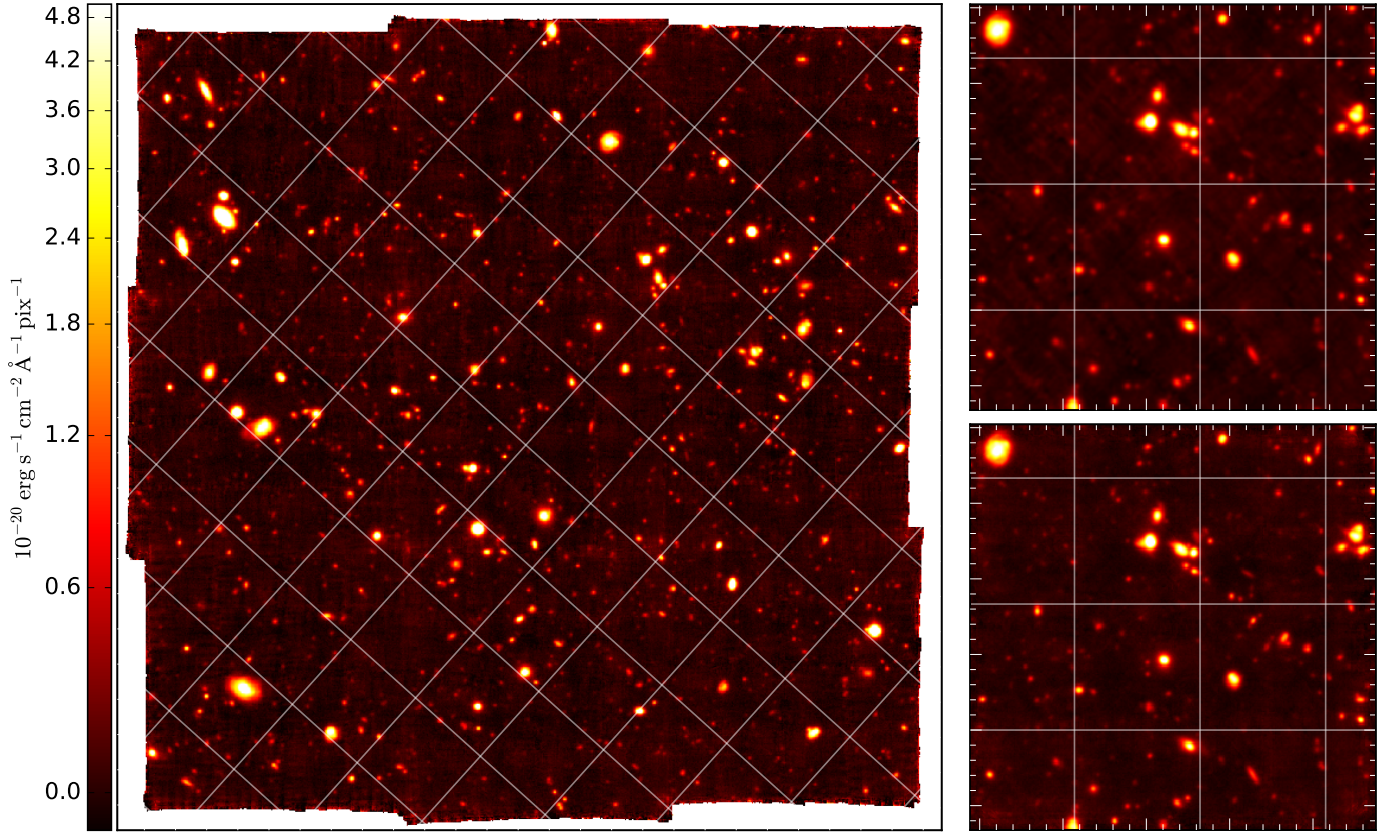
can also see some systematic offsets in the HDFS background at blue wavelengths which are not seen in the udf-10.

## 4. Astrometry and photometry

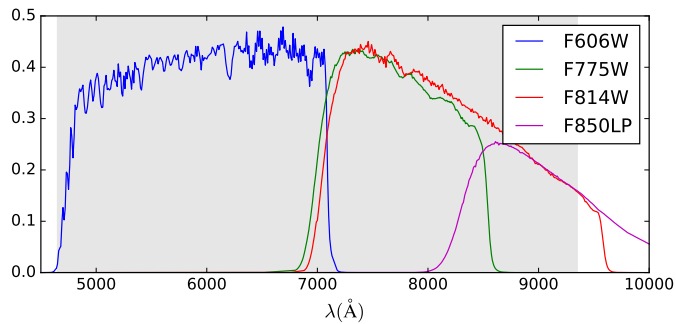
In the next sections we derive the broadband properties of the mosaic and udf-10 datacubes by comparing their astrometry and photometry to the HST broadband images.

We derive the MUSE equivalent broadband images by a simple weighted mean of the datacubes using the ACS/WFC filter response (Fig. 10). Note that the  $F606W$  and  $F775W$  filters are fully within the MUSE wavelength range, but the two others filters ( $F814W$  and  $F850LP$ ) extend slightly beyond the red limit. The corresponding HST images from the XDF data release (Illingworth et al. 2013) are then broadened by convolution to match the MUSE PSF (see Sect. 5.1) and the data are rebinned to the MUSE  $0''.2$  spatial sampling. For the comparison with the mosaic datacube, we split the HST images into the





**Fig. 9.** Reconstructed white light images for the mosaic (PA =  $-42^\circ$ , *left panel*) and the udf-10 (PA =  $0^\circ$ , *bottom right panel*). The mosaic rotated and zoomed to the udf-10 field is shown for comparison in the *top right panel*. The grid is oriented (north up, east left) with a spacing of  $20''$ .

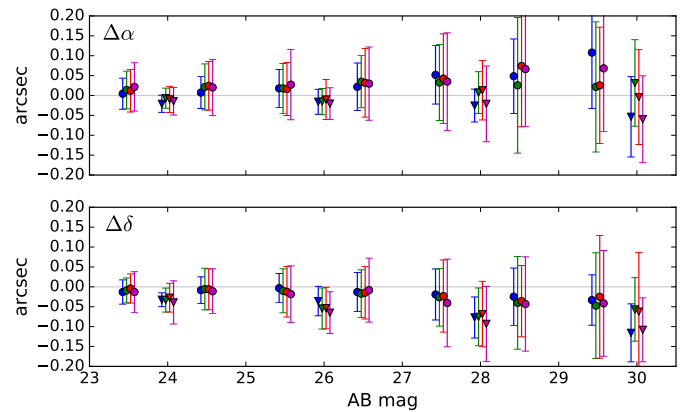


**Fig. 10.** ACS/WFC HST broadband filter response. The gray area indicates the MUSE wavelength range.

corresponding nine MUSE sub-fields in order to use the specific MUSE PSF model for each field.

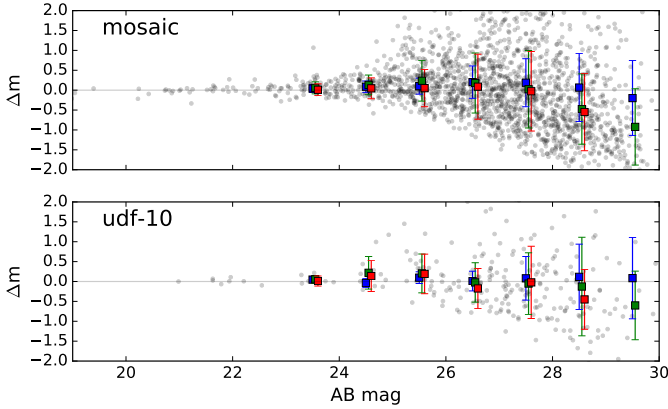
#### 4.1. Astrometry

The NoiseChisel software (Akhlaghi & Ichikawa 2015) is used to build a segmentation map for each MUSE image. NoiseChisel is a noise-based non-parametric technique for detecting nebulous objects in deep images and can be considered as an alternative to SExtractor (Bertin & Arnouts 1996). NoiseChisel defines “clumps” of detected pixels which are aggregated into a segmentation map. The light-weighted centroid is computed for each object and compared to the light-weighted centroid derived from the PSF-matched HST broadband image using the same segmentation map.



**Fig. 11.** Mean astrometric errors in  $\alpha$ ,  $\delta$  and their standard deviation in HST magnitude bins. The error bars are color coded by HST filter: blue (*F606W*), green (*F775W*), red (*F814W*) and magenta (*F850LP*). The two different symbols (circle and arrow) identify respectively the mosaic and udf-10 fields. Note that mosaic data are binned in 1-mag steps while udf-10 data points are binned over 2-mag steps in order to get enough points for the statistics.

The results of this analysis are given in Fig. 11 for both fields and for the four HST filters. As expected, the astrometric precision is a function of the object magnitude. There are no major differences between the filters, except for a very small increase of the standard deviation of the reddest filters. For objects brighter than AB 27, the mean astrometric offset is less than  $0''.035$  in the mosaic and less than  $0''.030$  in the udf-10. The standard deviation increases with magnitude, from  $0''.04$  for bright objects



**Fig. 12.** Differences between MUSE and HST AB broadband magnitudes. The gray points show the individual measurements for the  $F775W$  filter. The mean AB photometric errors and their standard deviations in HST magnitude bins are shown as error bars, color coded by HST filter: blue ( $F606W$ ), green ( $F775W$ ) and red ( $F814W$ ). *Top and bottom panels* respectively show the mosaic and udf-10 fields.

up to  $0''.15$  at  $AB > 29$ . For galaxies brighter than  $AB\ 27$ , we achieve an astrometric precision better than  $0''.07$  rms, i.e., 10% of the spatial resolution.

#### 4.2. Photometry

We now compute the broadband photometric properties of our data set, using a process similar to the previous astrometric measurements. This time, however we use the NoiseChisel segmentation maps generated from the PSF-matched HST broadband images. The higher signal-to-noise of these HST images allows us to identify more (and fainter) sources than in the MUSE equivalent image. The magnitude is then derived by a simple sum over the apertures identified in the segmentation map. We note that the background subtraction was disabled in order to measure the offset in magnitude between the two images. The process is repeated on the MUSE image using the same segmentation map and the magnitude difference saved for analysis. Note also that we exclude the  $F850LP$  filter in this analysis because a significant fraction of its flux ( $\approx 20\%$ ) lies outside the MUSE wavelength range.

The result of this comparison is shown in Fig. 12. The MUSE magnitudes match their HST counterparts well, with little systematic offset up to  $AB\ 28$  ( $\Delta m < 0.2$ ). For fainter objects, MUSE tends to under-estimate the flux with an offset more prominent in the red filters. The exact reason for this offset is not known but it may be due to some systematic left over by the sky subtraction process. As expected, the standard deviation increases with magnitude and is larger in the red than in the blue, most probably because of sky residuals. For example, the mosaic scatter is 0.4 mag in  $F606W$  at 26.5 AB, but is a factor of two larger in the  $F775W$  and  $F814W$  filters at the same magnitude. By comparison, the deeper udf-10 datacube achieves better photometric performance with a measured rms that is 20–30% lower than in the mosaic.

### 5. Spatial and spectral resolution

A precise knowledge of the achieved spatial and spectral resolution is key for all subsequent analysis of the data. For ground based observations where the exposures are obtained under various, and generally poorly known, seeing conditions, knowledge

of the spatial PSF is also important for each individual exposure. For example, exposures with bad seeing will add more noise than signal for the smaller sources and should be discarded in the final combination of the exposures. Note that the assessment of the spatial PSF for each individual exposure does not need to be as precise as for the final combined datacube.

The spectral resolution is not impacted by the change of atmospheric conditions and the instrument is stable enough to avoid the need of a spectral PSF evaluation for each individual exposure. However, good knowledge of the spectral resolution in the final datacube is also required.

In the next sections we describe the results and the methods used to derive these PSFs. To distinguish between the spectral and spatial axes, we name the spectral line spread function and the field spatial point spread function, LSF and FSF, respectively.

#### 5.1. Spatial point spread function (FSF)

In the ideal case of a uniform FSF over the field of view, its evaluation is straightforward if one has a bright point source in the field. If we assume a Gaussian shape, then only one parameter, the FWHM, fully characterizes the FSF. In our case we are not far from this ideal case, because the MUSE field is quite small with respect to the telescope field of view and its image quality ( $\sim 0''.2$ ) is much better than the seeing size. However, given the long wavelength range of MUSE, one cannot neglect the wavelength dependence of the seeing. For the VLT’s large aperture, a good representation of the atmospheric turbulence is given by Tokovinin (2002) in the form of a finite outer scale von Karman turbulence model. It predicts a nearly linear decrease of FWHM with respect to wavelength, with the slope being a function of the atmospheric seeing and the outer scale turbulence.

During commissioning, a detailed analysis of the MUSE FSF showed that it was very well modeled by a Moffat circular function  $[1 - (r/\alpha)^\beta]^{-\frac{1}{2}}$  with  $\beta$  constant and a linear variation of  $\alpha$  with wavelength. The same parametrization was successfully used in the HDFs study (Bacon et al. 2015) using the brightest star ( $R = 19.6$ ) in the field. However, most of MUSE UDF fields do not have such a bright star and the majority of our fields have no star with  $R < 23$  at all.

Fortunately, broadband HST images of the UDF exist for many wavelengths. In particular, as shown in Fig. 10, the wavelength coverage of four HST imaging filters,  $F606W$ ,  $F775W$ ,  $F814W$  and  $F850LP$  falls entirely or partially within the MUSE wavelength range (4750–9350 Å). If one of these images is convolved with the MUSE FSF, and the equivalent MUSE image is convolved with the HST FSF, then the resulting images should end up with the same combined FSF. Thus, the similarity of HST and MUSE images that have been convolved with models of each other’s FSFs, can be used to determine how well those models match the data.

In the following equations, suffixes of m and h are used to distinguish between symbols associated with the MUSE and HST images, respectively. Equation (1) models a MUSE image ( $\mathbf{d}_m$ ) as a perfect image of field sources ( $s$ ) convolved with the MUSE FSF ( $\psi_m$ ), summed with an image of random noise ( $\mathbf{n}_m$ ). Equation (2) is the equivalent equation for an HST image of the same region of the sky, but this time convolved with the HST FSF ( $\psi_h$ ), and summed with a different instrumental noise image,  $\mathbf{n}_h$ .

$$\mathbf{d}_m = s \times \psi_m + \mathbf{n}_m, \quad (1)$$

$$\mathbf{d}_h = s \times \psi_h + \mathbf{n}_h. \quad (2)$$

When these images are convolved with estimated models of each other's FSF, the result is as follows:

$$\mathbf{d}_m \times \psi'_h = s \times \psi_m \times \psi'_h + \mathbf{n}_m \times \psi'_h, \quad (3)$$

$$\mathbf{d}_h \times \psi'_m = s \times \psi_h \times \psi'_m + \mathbf{n}_h \times \psi'_m. \quad (4)$$

In these equations,  $\psi'_m$  and  $\psi'_h$  denote models of the true MUSE and HST FSF profiles,  $\psi_m$  and  $\psi_h$ . The following equation shows the difference between these two equations;

$$\mathbf{d}_m \times \psi'_h - \mathbf{d}_h \times \psi'_m = s \times (\psi_m \times \psi'_h - \psi_h \times \psi'_m) + (\mathbf{n}_m \times \psi'_h - \mathbf{n}_h \times \psi'_m). \quad (5)$$

The magnitude of the first bracketed term can be minimized by finding accurate models of the MUSE and HST FSFs. However, this is not a unique solution, because the magnitude can also be minimized by choosing accurate models of the FSF profiles that have both been convolved by an arbitrary function. To unambiguously evaluate the accuracy of a given model of the MUSE FSF, it is thus necessary to first obtain a reliable independent estimate of the HST FSF. This can be achieved by fitting an FSF profile to bright stars within the wider HST UDF image.

Minimizing the first of the bracketed terms of Eq. (5) does not necessarily minimize the overall equation. The noise contribution from the second of the bracketed terms decreases steadily with increasing FSF width, because of the averaging effect of wider FSFs, so the best-fit MUSE FSF is generally slightly wider than the true MUSE FSF. However provided that the image contains sources that are brighter than the noise, the response of the first bracketed term to an FSF mismatch is greater than the decrease in the second term, so this bias is minimal.

In summary, with a reliable independent estimate of the HST FSF<sup>6</sup>, a good estimate of the MUSE FSF can be obtained by minimizing the magnitude of Eq. (5), as a function of the model parameters of the FSF. In practice, to apply this equation to digitized images, the pixels of the MUSE and HST images must sample the same positions on the sky, have the same flux calibration, and have the same spectral response. A MUSE image of the same spectral response as an HST image can be obtained by performing a weighted mean of the 2D spectral planes of a MUSE cube, after weighting each spectral plane by the integral of the HST filter curve over the bandpass of that plane.

HST images have higher spatial resolutions than MUSE images, so the HST image must be translated, rotated and down-sampled onto the coordinates of the MUSE pixel grid. Before down-sampling, a decimation filter must be applied to the HST image, both to avoid introducing aliasing artifacts, and to remove noise at high spatial frequencies, which would otherwise be folded to lower spatial frequencies and reduce the signal-to-noise ratio (S/N) of the downsampled image. The model of the HST FSF must then be modified to account for the widening effect of the combination of the decimation filter and the spatial frequency response of the widened pixels.

Once the HST image has been resampled onto the same pixel grid as the MUSE image, there are usually still some differences between the relative positions of features in the two images, due to derotator wobble and/or telescope pointing errors. Similarly, after the HST pixel values have been given the same flux units as the MUSE image, the absolute flux calibration factors and offsets of the two images are not precisely the same. To correct these residual errors, the MUSE FSF fitting process has to simultaneously fit for position corrections and calibration corrections, while also fitting for the parameters of the MUSE FSF.

<sup>6</sup> In practice we compute the Moffat fit for a few bright stars in the field for each HST filter.

The current fitting procedure does not attempt to correct for rotational errors in the telescope pointing, or account for focal plane distortions. Focal plane distortions appear to be minimal for the HST and MUSE images, and only two MUSE images were found to be slightly rotated relative to the HST images. In the two discrepant cases, the rotation was measured by hand, and corrected before the final fits were performed.

As described earlier, the FSF of a MUSE image is best modeled as a Moffat function. Moffat functions fall off relatively slowly away from their central cores, so a large convolution kernel is needed to accurately convolve an image with a MUSE FSF. Convolution in the image plane is very slow for large kernels, so it is more efficient to perform FSF convolutions in the Fourier domain. Similarly, correcting the pointing of an image by a fractional number of pixels in the image domain requires interpolation between pixels, which is slow and changes the FSF that is being measured. In the Fourier domain, the same pointing corrections can be applied quickly without interpolation, using the Fourier-transform shift theorem. For these reasons, the FSF fitting process is better performed entirely within the Fourier domain, as described below.

Let  $b$  and  $\gamma$  be the offset and scale factor needed to match the HST image photometry to that of the MUSE image, and let  $\epsilon$  represent the vector pointing-offset between the HST image and the MUSE image. When the left side of Eq. (5) is augmented to include these corrections, the result is the left side of the following equation:

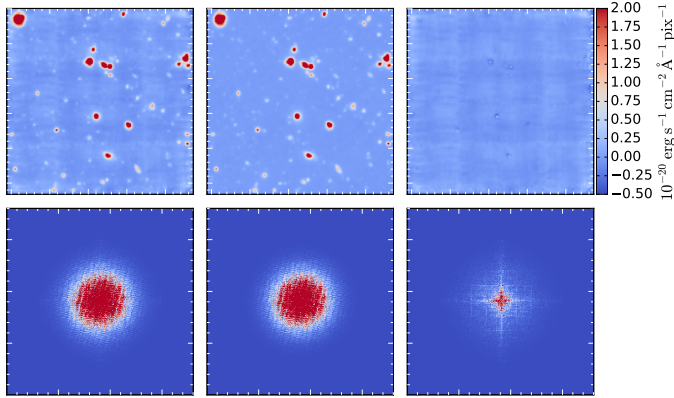
$$\mathbf{d}_m \times \psi'_h - \gamma \mathbf{d}_h \times \psi'_m \times \Delta(\mathbf{p} - \epsilon) + b \xrightarrow{\text{FT}} \mathbf{D}_m \Psi'_h - \gamma \mathbf{D}_h \Psi'_m e^{-i2\pi f \epsilon} + b. \quad (6)$$

Note that the pointing correction vector ( $\epsilon$ ) is applied by convolving the HST image by the shifted Dirac delta function,  $\Delta(\mathbf{p} - \epsilon)$ , where  $\mathbf{p}$  represents the array of pixel positions.

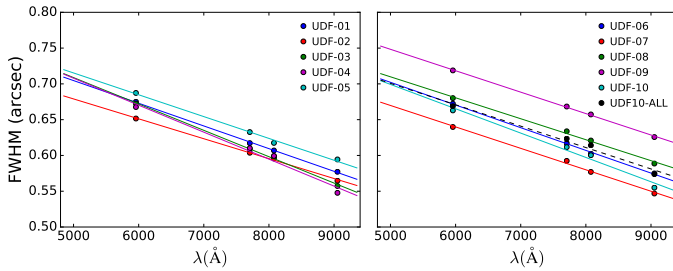
The right side of Eq. (6) is the Fourier transform of the left side, with  $\mathbf{d}_h \xrightarrow{\text{FT}} \mathbf{D}_h$ ,  $\mathbf{d}_m \xrightarrow{\text{FT}} \mathbf{D}_m$ ,  $\psi_m \xrightarrow{\text{FT}} \Psi_m$  and  $\psi_h \xrightarrow{\text{FT}} \Psi_h$ . The spatial frequency coordinates of the Fourier transform pixels are denoted  $f$ . Note that all of the convolutions on the left side of the equation become simple multiplications in the Fourier domain. The exponential term results from the Fourier transform shift theorem, which, as shown above, is equivalent to an image-plane convolution with a shifted delta function.

The fitting procedure uses the Levenberg-Marquardt non-linear least-squares method to minimize the sum of the squares of the right side of Eq. (6). The procedure starts by obtaining the discrete Fourier transforms,  $\mathbf{D}_m$ ,  $\mathbf{D}_h$ , and  $\Psi'_h$  using the Fast Fourier Transform (FFT) algorithm. Then for each iteration of the fit, new trial values are chosen for  $\gamma$ ,  $b$ ,  $\epsilon$  and the model parameters of the MUSE FSF,  $\psi_m$ . There is no analytic form for the Fourier transform of a 2D Moffat function, so at each iteration of the fit, the trial MUSE FSF must be sampled in the image plane, then transformed to the Fourier domain using an FFT. It is important to note that to avoid significant circular convolution, all images that are passed to the FFT algorithm should be zero padded to add margins that are at least as wide as the core of the trial Moffat profiles and the maximum expected pointing correction.

MUSE and HST images commonly contain pixels that have been masked due to instrumental problems, or incomplete field coverage. In addition, areas of the images that contain nearby bright stars should be masked before the FSF procedure, because the effect of the proper motion of these stars is often sufficiently large between the epochs of MUSE and HST observations, to



**Fig. 13.** An example demonstrating the success of the FSF fitting technique. The *upper left panel* shows the udf-10 data, rescaled by the equivalent HST *F775W* broadband filter. The *upper middle panel* shows the corresponding HST *F775W* image, after it has been resampled onto the pixel grid of the MUSE image and convolved with the best-fit MUSE FSF. The *upper right panel* presents the residual of these two images, showing that only the instrumental background of the MUSE image remains. The *lower panels* show the corresponding images in the Fourier space where the fit is performed.



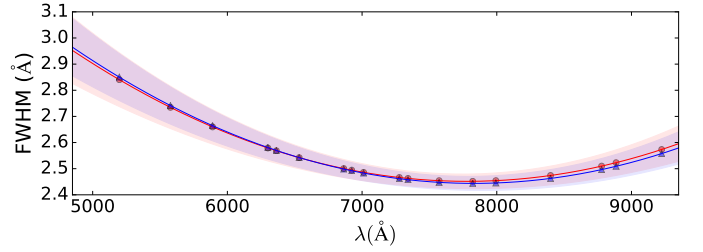
**Fig. 14.** FSF Fitting results for all mosaic and udf-10 fields. For each field, 4 fit MOFFAT FWHMs corresponding to 4 HST filters (*F606W*, *F775W*, *F814W*, *F850LP*) are displayed, together with the linear fit. The UDF10-ALL is for the combined depth of the udf-10 field and its associated mosaic fields (1, 2, 4 and 5).

make it impossible to line up the stars without misaligning other sources. Since the FFT algorithm cannot cope with missing samples, masked pixels must be replaced by a finite value. Here, we choose a replacement value of zero, since this choice makes the fit of the calibration scale factor ( $\gamma$ ) insensitive to the existence of missing pixels. However, a contiguous region of zero-valued pixels can fool the algorithm, making it think the region (which is significantly different from its surroundings) is a real feature to be fit. To avoid this, we first subtract the median flux value from each image before replacing the masked pixels with zero. This decreases the contrast around the masked pixels, increasing the probability that they will blend into the background and be ignored by the fitting routine. The median-subtracted flux value is saved and folded into the fit of the background offset parameter ( $b$ ).

Figure 13 shows an example of how well this method works in practice and Fig. 14 displays the fitting results obtained for all fields. The fit values for the combined datacubes of each field are given in Table 1.

## 5.2. Spectral line spread function (LSF)

To measure the LSF, we produce combined datacubes similar to the udf-10 and mosaic datacubes but without including the sky



**Fig. 15.** Measured mean LSF FWHM on the udf-10 (blue line) and mosaic (red line) datacubes. The symbols represent measured values while the solid line represents the polynomial fit. The shaded area shows the  $\pm 1\sigma$  spatial standard deviation.

subtraction. From these, we calculate the LSF using 19 groups of 1–10 sky lines. While the lines within each group are unresolved at the MUSE spectral resolution, they must be accounted for to construct a proper LSF model. For each group we used the CAMEL software (see [Epinat et al. 2012](#); [Contini et al. 2016](#), for a description of the software) to fit a Gaussian to each line, keeping the relative position and FWHM identical for all lines in the group. This is performed over all spaxels in the datacube, after applying a Gaussian spatial smoothing kernel of  $0''.4$  FWHM to improve the S/N of the faint sky lines.

We show the mean and standard deviation of the resulting FWHM as a function of wavelength in Fig. 15. Note that there is, as expected, little difference between the udf-10 and mosaic datacubes. The FWHM of the modeled LSF varies smoothly with wavelength, ranging from  $3.0 \text{ \AA}$  (at the blue end) to  $2.4 \text{ \AA}$  (at  $7500 \text{ \AA}$ ). It remains largely constant over the field of view, with an average standard deviation of  $0.05 \text{ \AA}$ . The FWHM variations as a function of wavelength  $F(\lambda)$  (in  $\text{\AA}$ ) are best described by polynomial functions:

$$F_{\text{mosaic}}(\lambda) = 5.835 \times 10^{-8} \lambda^2 - 9.080 \times 10^{-4} \lambda + 5.983 \quad (7)$$

$$F_{\text{udf10}}(\lambda) = 5.866 \times 10^{-8} \lambda^2 - 9.187 \times 10^{-4} \lambda + 6.040. \quad (8)$$

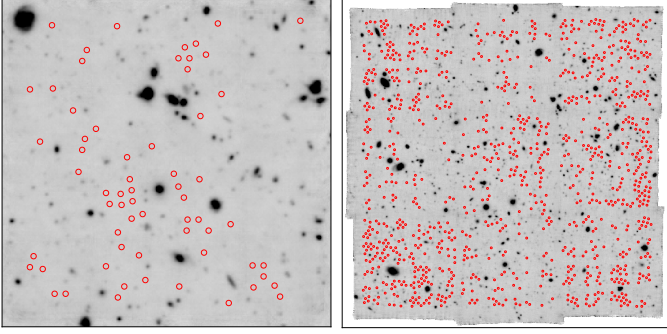
We note that the true LSF shape is not actually Gaussian, but instead more square in shape. The simple Gaussian model is however a good approximation for most usage.

## 6. Noise properties and limiting flux

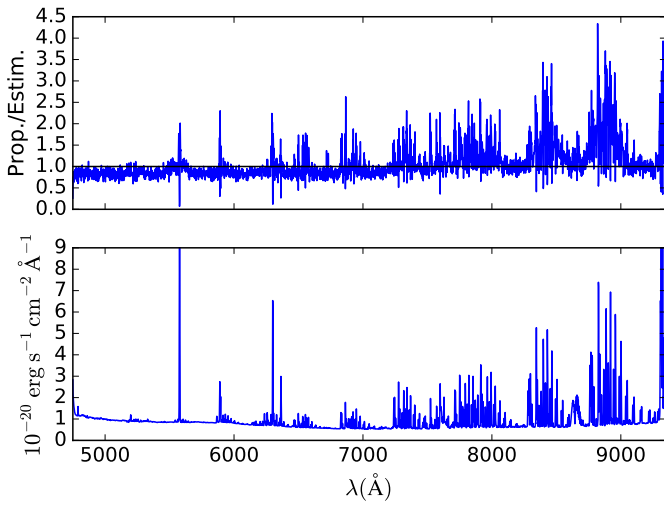
### 6.1. Noise properties

The empirical procedure described in Sect. 3.1.5 should correct the variance estimate for the correlation added by the 3D drizzle interpolation process. We thus expect the propagated variance of the final datacubes to be correct in that respect. To check that this is indeed the case, we estimate the variance from a set of empty regions in the datacubes, selected to have similar integration time using the exposure maps shown in Fig. 2. For the udf-10 field, we select 63 circular apertures of  $1''$  diameter in regions with  $31 \pm 0.3$  h of integration time. In the mosaic we select 991 similar apertures in regions with  $9.9 \pm 0.4$  h of integration time. The locations of all selected regions are shown in Fig. 16.

We calculate the corresponding propagated variance spectrum by taking the median of the stack of all apertures. The spectrum generated from the udf-10 field, along with the ratio between this standard deviation and the estimated standard deviation calculated in Sect. 3.1.5 are shown in Fig. 17. As expected,



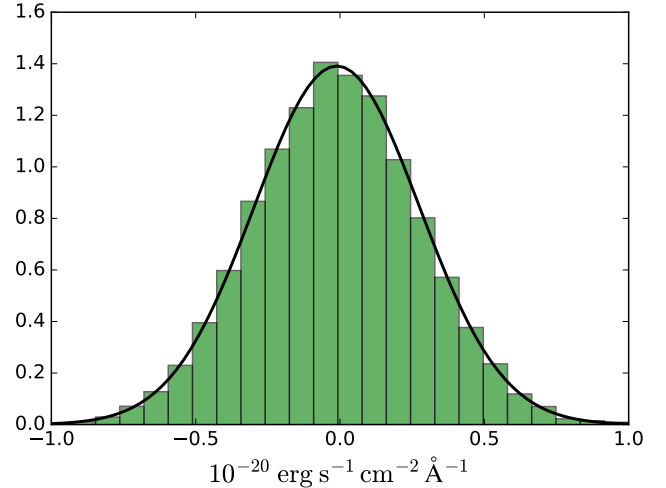
**Fig. 16.** Selected apertures used to evaluate the variance in empty regions of the udf-10 (*left panel*) and mosaic (*right panel*) datacubes.



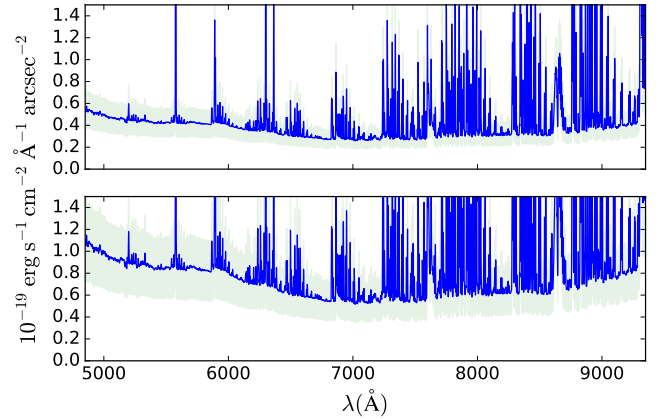
**Fig. 17.** *Lower panel:* median value of the propagated noise standard deviation for the 63 selected 1'' diameter apertures (see text). *Top panel:* ratio of the propagated to the estimated standard deviations.

the computed ratio is around unity<sup>7</sup> and constant with wavelength, showing that the propagated variance is now a good representation of the true variance within an aperture. In the top panel of Fig. 17 it is clear that there is a mismatch between estimated and propagated standard deviation at wavelengths that contain bright sky emission lines. The difference is due to the PCA ZAP process and discussed in detail in Sect. 5 of Soto et al. (2016): when ZAP is applied to the individual datacubes (see Sect. 3.1.4) it tends to preferentially remove the strongest systematic signals left by the imperfect sky subtraction at the locations of the bright sky lines. For the brightest OH lines this results in an over-fitting of the noise which then biases the estimated variance. In that respect, the propagated variance is a better representation of the true variance. The same behavior is found for the mosaic datacube.

Using the set of empty apertures we are also able to investigate the noise probability density distribution. A normal test (Pearson et al. 1977) returns a  $p$ -value of  $\approx 0.3$ , demonstrating that the noise probability density distribution is normal with a high probability (see the example in Fig. 18).



**Fig. 18.** Example of a normalized data histogram derived from an empty aperture of 1'' diameter at 7125 Å in the udf-10 datacube. The solid line displays the best fit Normal PDF with a standard deviation of  $0.33 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1}$ .



**Fig. 19.**  $1\sigma$  surface brightness limit for the mosaic (*bottom*) and udf-10 (*top*) datacubes computed for an aperture of  $1'' \times 1''$ . The blue curve displays the average value and the green area the rms over the field of view.

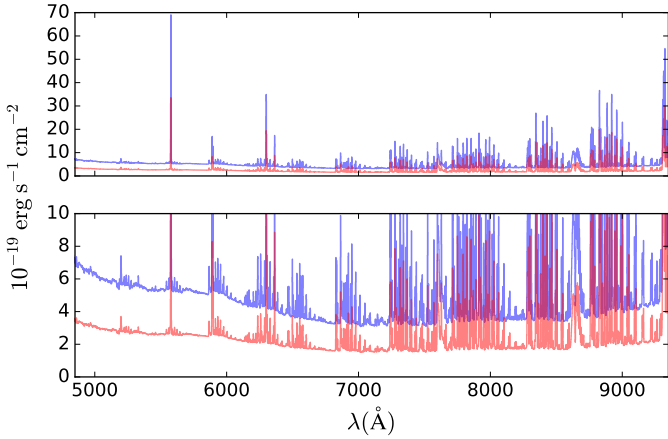
## 6.2. Limiting line flux

From the noise properties one can derive the limiting line flux. We start to evaluate the  $1\sigma$  emission line surface brightness limit by computing the (sigma-clipped) mean and standard deviation of the propagated variance over the complete field of view for the udf-10 and mosaic datacubes. The resulting emission line surface brightness limit is shown in Fig. 19. A  $1\sigma$  emission line sensitivity of  $2.8$  and  $5.5 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1} \text{ arcsec}^{-2}$  for an aperture of  $1'' \times 1''$  is reached in the  $7000\text{--}8500 \text{ Å}$  range for the udf-10 and mosaic datacubes, respectively.

Note that the measured value in the udf-10 ( $2.8$ ) is slightly better than what we would have predicted from the mosaic value ( $3.2$ ), taking into account the  $\sqrt{3}$  factor predicted by the difference in integration time. It shows that the observational strategy used for the udf-10 (see Sect. 2) is effective in further reducing the systematics which are still present in the mosaic datacube.

This result compares advantageously with the early HDFS observations (Bacon et al. 2015) which reached a  $1\sigma$  emission line surface brightness limit of  $4.5 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1} \text{ arcsec}^{-2}$  in the same aperture. The

<sup>7</sup> According to Fig. 17 the propagated standard deviation underestimate the compute values by  $\approx 10\text{--}15\%$  but we did not attempt to correct for this small offset.



**Fig. 20.**  $3\sigma$  emission line flux detection limit for point-like sources for the mosaic at 10 h integration time (in blue) and udf-10 at 31 h integration time (in red) datacubes. The full scale sky lines dominated limiting flux is shown in the *top panel*, while values outside bright sky lines are shown in the *bottom panel*.

1.6 better sensitivity<sup>8</sup> achieved with the udf-10 datacube is the result of the extensive work performed on observational strategy and data reduction since 2014. While the performance of the first release of the HDFs datacube was dominated by systematics, we have pushed the UDF datacubes to another level of quality and sensitivity.

We now derive the line flux detection limit for a point-like source, using weighted FSF extraction and summation over three spectral channels (i.e.,  $3.75 \text{ \AA}$ ). This value is of course dependent on the integration time (see the exposure map in Fig. 2). We give the  $3\sigma$  limiting line flux in Fig. 20 for the corresponding median integration times of the mosaic and udf-10 datacubes. The corresponding detection limits are  $1.5 \times 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2}$  and  $3.1 \times 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2}$  in the region around  $7000 \text{ \AA}$  between OH sky lines, for the udf-10 and mosaic datacubes, respectively.

## 7. Source detection and extraction

Exploration of the mosaic and udf-10 datacubes starts by finding sources, extracting their spatial and spectral information (e.g., subimages and spectra) and measuring their redshifts. The last step is discussed in Paper II (Inami et al. 2017). Here we discuss the first steps using two techniques: optimal source extraction with an HST prior, and blind detection of emission line objects.

### 7.1. HST-prior extraction

As an input to our HST-prior extraction, we use the locations of objects in the Rafelski et al. (2015) source catalog. This catalog provides precise astrometry, photometry and photometric redshifts for 9927 sources covering the entire UDF region.

Given the MUSE spatial resolution,  $0''.7$  versus  $0''.1$  for HST, our data are unfortunately impacted by source confusion. Thus, from the initial catalog, we compile a new catalog of 6288 sources, created by merging all Rafelski et al. (2015) sources which have a separation less than  $0''.6$ . For these merged systems, we compute a new source location based on the  $F775W$ -light-weighted centroid of all objects that make up the new merged source.

<sup>8</sup> This factor is probably a lower limit given that the noise analysis performed on the HDFs datacube did not fully take into account the correlated noise.

We then proceed to source extraction. Using the Rafelski et al. (2015) segmentation map, we extract each source from the MUSE data in a region defined by its original segmentation area convolved with a Gaussian of  $0''.6$  FWHM to take into account the MUSE resolution. We generate a series of 1D spectra from each extraction region, using several different weighting schemes: (a) a uniformly weighted, direct summation over the full segmentation area, (b) an optimally weighted sum using the reconstructed MUSE white-light image as the weight, and (c) an optimally weighted sum using the estimated FSF at the source location<sup>9</sup>. We also compute a second set of three spectra, using the same weighting schemes, after subtracting a background spectrum from the data. This spectrum is computed as the average over the empty region free of sources surrounding the object, using the convolved segmentation image as a guide.

The optimal extraction is based on the Horne (1986) algorithm;

$$f(\lambda) = \frac{\sum_x M_x W_{x,\lambda} (D_{x,\lambda} - S_\lambda) / V_{x,\lambda}}{\sum_x M_x W_{x,\lambda}^2 / V_{x,\lambda}} \quad (9)$$

$$v_f(\lambda) = \frac{\sum_x M_x W_{x,\lambda}}{\sum_x M_x W_{x,\lambda}^2 / V_{x,\lambda}} \quad (10)$$

where  $f(\lambda)$  is the optimal flux and  $v_f(\lambda)$  its variance,  $D$ ,  $S$  and  $V$  the data, sky and variance datacubes,  $M$  the segmentation mask and  $W$  the weight which is either the white-light image or the FSF. Depending on the object, one of these weighting schemes provides a higher S/N than the others. In general we use the background-subtracted white-light weighted spectra for bright and extended objects ( $AB < 26$  and  $FWHM > 0.5 \times FSF$ ) and background-subtracted FSF weighted spectra for other small and/or faint objects. An example of source extraction is shown in Fig. 21.

Due to the convolution, the segmentation map of one source can overlap with other neighboring sources, creating some blending effects in the extracted spectrum. In a number of cases, as shown in Paper II (Inami et al. 2017), the source can be *de-blended* using the reconstructed narrow-band location when an emission line is present. One such case can be seen in Fig. 21. In that figure, the three central HST sources cannot be resolved in the MUSE white light image and thus were originally merged into one source in the extraction process. However, the reconstructed narrow band image shows that the  $z = 4.1$  Ly $\alpha$  emission<sup>10</sup> can be clearly attributed to a unique HST object. Note that this galaxy forms a pair with another Ly $\alpha$  emitter (ID 412) at the same redshift located  $3''.5$  SE with respect to the source center.

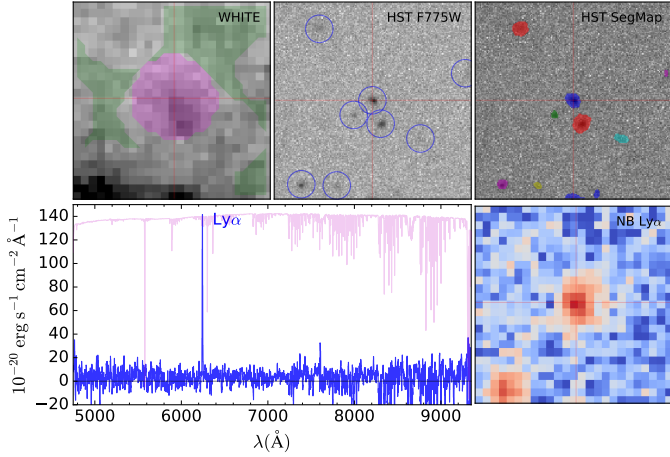
The extraction process is run independently for the mosaic and udf-10 datacubes, using the same input catalog in each case to ensure that objects which are in both datacubes receive the same ID.

### 7.2. Blind detection with ORIGIN

The HDFs study (Bacon et al. 2015) has demonstrated MUSE's ability to detect emission line galaxies without an HST counterpart, so we should not have to rely only on HST-prior source detection when searching for high equivalent-width star-forming galaxies in the UDF. Note however, that the HST data set covering the UDF reaches a  $5\sigma$  depth of 29.5 in the  $F775W$  filter, i.e.,

<sup>9</sup> In the case of overlapping fields, the FSF is computed as the average of all fields at the source location, weighted by the exposure map.

<sup>10</sup> The Ly $\alpha$  line was identified from its asymmetric profile and fainter continuum on the bluer side of the line.



**Fig. 21.** Source ID 6698 from the udf-10 datacube. On top, from left to right, one can see the MUSE reconstructed white light image, the HST image in  $F775W$  and the HST Rafelski segmentation map. Image size is  $5''$  and the source center is indicated by a red crosshair. The blue circles mark the sources identified in the Rafelski catalog. The central Rafelski source ID is 4451 and its  $F775W$  AB magnitude is  $27.92 \pm 0.04$ . The source and background masks are overlaid on the MUSE white light image in magenta and green colors, respectively. Bottom left: PSF weighted extracted source spectrum over the whole wavelength range (box-filtered with a window of 5 pixels). The noise standard deviation is shown in magenta (mirrored with respect to the source spectra). Bottom right:  $\text{Ly}\alpha$  Narrow-Band image.

one magnitude deeper than the HST HDFS observations. Therefore, we expect to find fewer sources without HST counterparts in the UDF, though this number is surely greater than zero. Because of this, it is beneficial to attempt to locate these “hidden” galaxies through the use of a blind detection algorithm.

Aside from looking for a specific class of galaxy, there is also a practical motivation for performing a blind search of the MUSE datacube. As discussed in Paper II, redshift assessment is a difficult task which (as of now) is not fully automated, instead relying in large part on expert judgement. In that respect, investigating all 9927 objects in the Rafelski et al. (2015) catalog is a tedious undertaking. However, the task can be alleviated by a blind search, assuming it can efficiently pre-select emission line objects.

Several tools have already been developed to perform blind searches of faint emitters in MUSE datacube, such as: MUSELET, a SExtractor based method available in MPDAF<sup>11</sup> (Piqueras et al. 2017), LSDCAT, a matching filter method (Herenz & Wisotzki 2017), SELFI, a Bayesian method (Meillier et al. 2016) and CubExtractor (Cantalupo, in prep.), a three-dimensional automatic extraction software based on connecting-labeling-component algorithm (used, e.g., in Borisova et al. 2016; and Fumagalli et al. 2016).

Each of these methods has its own pros and cons: some achieved high sensitivity but at the expense of low purity, others are optimized to provide reliable results, that is high purity, but with lower sensitivity. Given the depth and the field of view of the UDF observations, we expect to find thousands of emission line galaxies which, considering the MUSE spatial resolution, will include a significant fraction of blended sources. The total size of the datacube (3.3 billion voxels for the mosaic) is not negligible either. In order to handle these methodological and

computational challenges, we have begun to develop a new automated method, called ORIGIN.

The method is still in development and will be presented in a future paper (Mary et al., in prep.), but it is already mature enough to be efficiently used for the UDF blind search. In the following sections we briefly explain how the method works and show the results obtained for our observations.

### 7.2.1. Method

The basic idea of the algorithm is to follow a matched filter approach, where the filters are spatio-spectral (3D) signatures formed by a set of spectral templates (or profiles) that are spatially extended by the point spread function of the instrument (Paris et al. 2013). In practice, this approach alone is neither robust nor reliable, because the corresponding test statistic is highly sensitive to sources different from the ones of interest and to residual artifacts (both referred to as unknown nuisance signals). A standard approach in this situation is to model and estimate the nuisance signals under both hypotheses ( $\mathcal{H}_0$ : line absent;  $\mathcal{H}_1$ : line present), see for instance Kay (1998), Scharf & Friedlander (1994). However, the resulting tests are computationally intensive and seem hardly compatible with the datacube size. ORIGIN consequently opts for a two-step strategy, where the nuisance signals are suppressed first (using a standard Principal Component Analysis, hereafter PCA) and the lines are detected in the PCA residuals. The resulting test statistics are used to assign a probability to each predicted line. For each line that is flagged as significant, a narrow band (NB) test is performed in order to check whether the line is also significant in the raw data, that is, before any processing (weighting by the estimated variances, PCA) is performed. This step is required because variance underestimation (especially around sky lines, see Sect. 3.1.5) may create artificial lines when weighting the data. Each line that survives the NB test is estimated (deconvolved), leading to an estimate of the line center (a triplet of two spatial and one spectral coordinates). The lines are then merged into sources, leading to a catalog of sources with estimated lines and various other information.

**Suppression of nuisance signals:** To be consistent with a likelihood based approach, the whole datacube is first weighted by the estimated standard deviation of the noise in each voxel (Sect. 3.1.5). In order to account for spatially varying statistics (regions with more or less bright and/or extended sources) the cube is segmented spatially into several regions (16 for the udf-10 and 121 for the mosaic). For a given region, each std-weighted data pixel  $\mathbf{p}$  (a vector whose length is the number of spectral channels) is modeled as a continuum  $\mathbf{c}$  plus a residual  $\mathbf{r}$ :  $\mathbf{p} = \mathbf{c} + \mathbf{r}$ . The continuum is assumed to belong to a low dimensional subspace, which is obtained by a PCA of all pixels of the considered region. The number of eigenvectors spanning this subspace is computed adaptively for each region. If  $\mathbf{V}_z$  denotes the matrix of the retained (orthonormal) eigenvectors, the residual is estimated as  $\hat{\mathbf{r}} = \mathbf{p} - \hat{\mathbf{c}} = \mathbf{p} - \mathbf{V}\mathbf{V}_z^T\mathbf{p}$ . This analysis produces a cube of residuals and, as a side product, a cube of continuum spectra.

**Line search:** For all angular and spectral positions  $(\alpha, \delta, \lambda)$  in the residual datacube, the line search considers subcubes of the size of the considered target signatures (typically  $13 \text{ px} \times 13 \text{ px} \times 20$  spectral channels, representing  $2''.6 \times 2''.6 \times 25 \text{ \AA}$ ) and makes, for each subcube  $s$  centered at location  $(\alpha_s, \delta_s, \lambda_s)$ , a

<sup>11</sup> See the MPDAF MUSELET documentation at <http://mpdaf.readthedocs.io/en/latest/muselet.html>

test for the two hypotheses:

$$\begin{cases} \mathcal{H}_0: \mathbf{s} = \mathbf{n} & \text{(Noise only),} \\ \mathcal{H}_1: \mathbf{s} = \alpha \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{d} + \mathbf{n} & \text{(Line centered at } (x_s, y_s, \lambda_s) \text{ plus noise),} \end{cases}$$

where  $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  is the noise – assumed to be a zero-mean Gaussian with an Identity covariance matrix,  $\mathbf{\Sigma}$  denotes the noise covariance matrix of the data before weighting (assumed to be diagonal in absence of information on noise correlations),  $\alpha > 0$  is the unknown amplitude of the emission line and  $\mathbf{d}$  is a spatio-spectral profile weighted by the local values of the noise standard deviation (the weights change for each tested voxel  $x_s, y_s, \lambda_s$ ). The profile  $\mathbf{d}$  is unknown but assumed to belong to a dictionary of 12 spectral profiles (say,  $\mathbf{d}_i, i = 1, \dots, 12$ ) of various widths (from 3 to 16 Å) convolved by the local (wavelength dependent) FSF (a Moffat function). A Generalized Likelihood Ratio (GLR) approach leads to a test statistic  $T(\mathbf{s})$  in the form of a weighted correlation:

$$T(\mathbf{s}) = \max_i \frac{\mathbf{s}^\top \mathbf{\Sigma}^{-1} \mathbf{d}_i}{\|\mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{d}_i\|},$$

for which the numerator and denominator can be efficiently computed using fast convolutions.

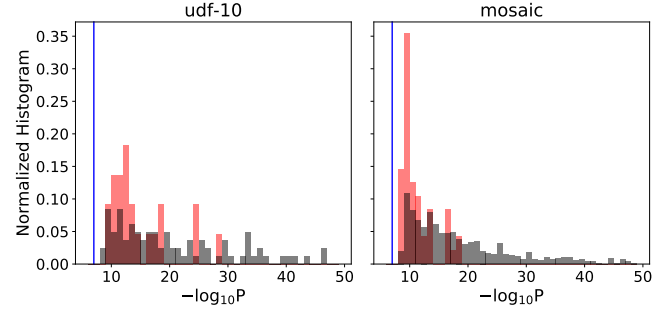
***P*-values:** For the correlations, the *P*-value associated to an observed correlation  $t$  is:  $p_T := \Pr(T > t \mid H_0)$ . A *P*-value measures how unlikely a test statistic is under the null hypothesis. The distribution of  $T$  under the null hypothesis is estimated from the data in each region and *P*-values  $p_T$  are computed for each voxel position. All voxels with a *P*-value below a threshold (set after some trial to  $10^{-7}$  for the UDF datacubes, corresponding to a detection limit of  $5.2\sigma$  for a Gaussian PDF) are flagged as significant.

In the current version, the algorithm also computes a probability that each spectral channel is not contaminated by residual artifacts (such as spurious residuals from sky lines subtraction), by comparing the number of significant *P*-values in each channel against what should be expected from a uniform distribution of noise.

The final probabilities (in the form of *P*-values) evaluate the probability that the line is significant at each voxel position conditionally to the fact that the considered voxel does not belong to a channel contaminated by artifacts. The *P*-values less than a threshold (set to  $10^{-7}$ ) survive this step.

Thresholding the *P*-values leads to clusters of significant *P*-values, because the signature of a line generally leads to several small *P*-values located in a group of voxels in the vicinity of the line center. To determine a first estimate of the position of the line center, the algorithm retains the smallest *P*-value in each group.

**Narrow band tests:** For each detected line, this step defines a subcube  $\mathbf{t}$  in the raw data centered on the supposed line location and a control subcube  $\mathbf{b}$  further away in wavelength (3 times the spectral length of the profiles that created the detection, say  $\mathbf{d}_k$ ). A GLR test is then conducted between two hypotheses: under the null hypothesis, both subcubes contain an unknown constant background plus noise, and under the alternative hypothesis  $\mathbf{t}$  also contains the line  $\mathbf{d}_k$  with an unknown amplitude. The test keeps all lines for which the test statistic  $\frac{(\mathbf{t}-\mathbf{b})^\top \mathbf{d}_k}{\sqrt{2}\|\mathbf{d}_k\|}$  is larger than a threshold (set to 2 for the UDF).



**Fig. 22.** Normalized histograms of the *P*-values of the ORIGIN sources with (in gray, restricted to Ly $\alpha$  emitters) and without (in red) HST counterpart. The blue line displays the threshold *P*-value ( $10^{-7}$ ).

**Line estimation:** The spectral profile and spatial position of each detected line is estimated by spatial deconvolution. The final spectral position of the line is the maximum of the estimated line. Note that while Gaussian profiles are used for detecting the line, this step allows for the recovery of any line profile, for instance asymmetric or double lines.

**Catalog output:** The lines are merged into sources by moving over the angular coordinates of the cube containing all detected line centers within a cylinder of diameter equal to the FWHM of the FSF (averaged over the spectral channels) and  $z$  axis aligned with the spectral axis. For each object, the algorithm outputs are an ID number, its angular position and the detected lines. The spectral channel of the line, Gaussian profile that created the initial detection, correlation and spectral channel tests' *P*-values, NB test scores, NB images, deconvolved line profile, estimated flux and FWHM are stored for each line.

### 7.2.2. Application to the UDF

The ORIGIN algorithm is implemented as a Python package and was successfully run on the UDF fields using the parameters defined in the previous section. The full computation takes  $\approx 1$  and  $\approx 6$  h of computing time on our 32 multi-core linux workstation for udf-10 and mosaic datacubes, respectively. The program reported the detection of 355 (udf-10) and 1923 (mosaic) candidate sources. After removing the 49 (udf-10) and 672 (mosaic) false detections<sup>12</sup> identified after visual inspection, we are left with 306 and 1251 potentially *real* detections, corresponding to 86% and 65% purity, respectively, for the udf-10 and mosaic fields.

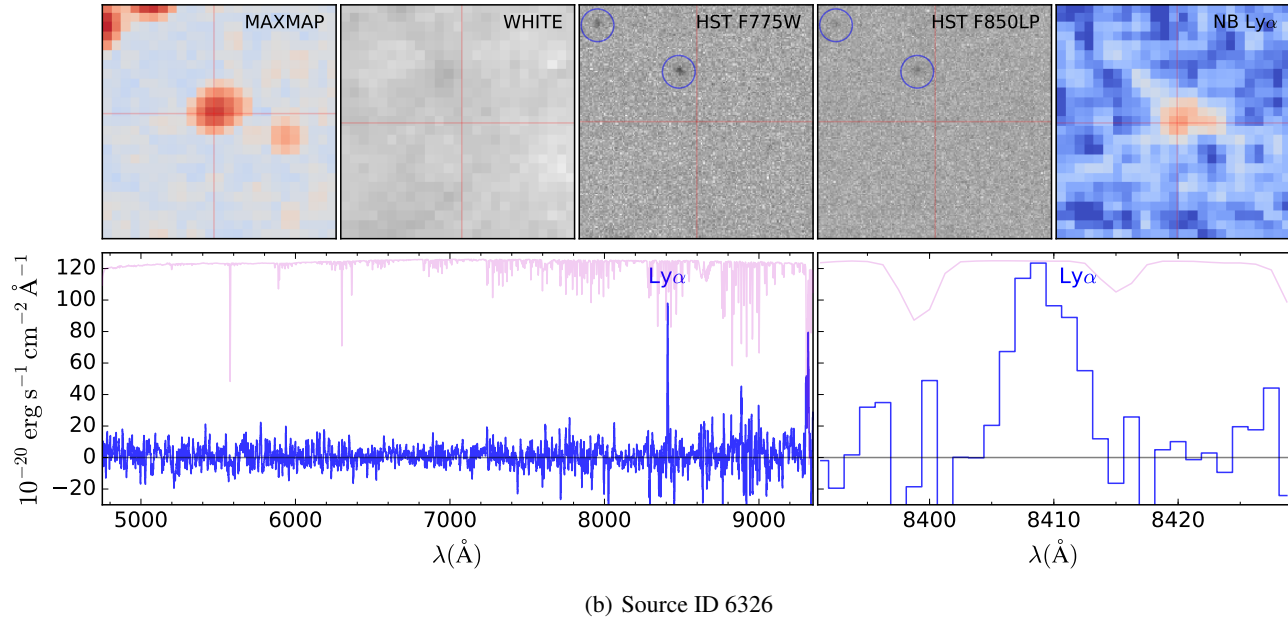
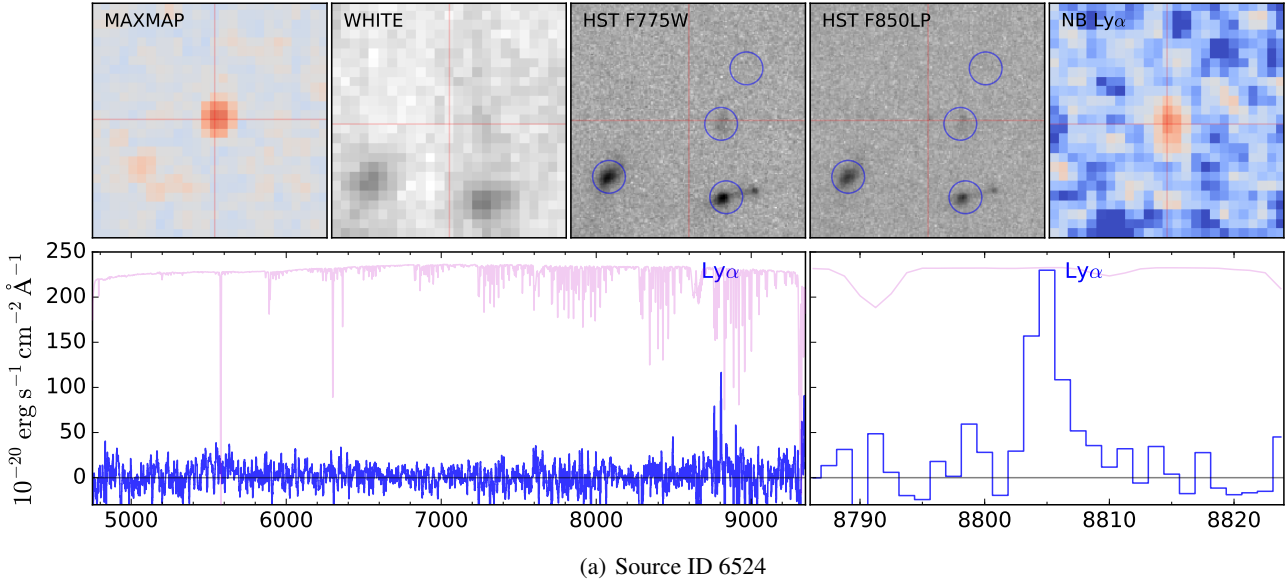
As shown in Paper II (Inami et al. 2017), not all detections will eventually turn into a redshift. Generally, the detected sources without redshift have a S/N that is too low to identify the emission and/or absorption lines, but the vast majority at least have an HST counterpart, validating their detection status.

A comparative analysis between the ORIGIN-detected and the HST-prior extracted sources is presented in Paper II. This comparison has been fruitful in finding the remaining problems with ORIGIN which impact its sensitivity and/or its purity, which will result in an improved version in the near future. However, despite its current limitations, ORIGIN is able to detect a large number of sources, especially high-redshift, faint Ly $\alpha$  emitters.

One such example is given in Fig. 23a. The source is detected at high significance by ORIGIN ( $P < 10^{-9}$ ) in the mosaic datacube, as can be seen in the MAXMAP image. This

<sup>12</sup> The false detections are mainly due to residuals left over by continuum subtraction, splitting of extended bright sources in multiple sources plus a few remaining datacube defects.





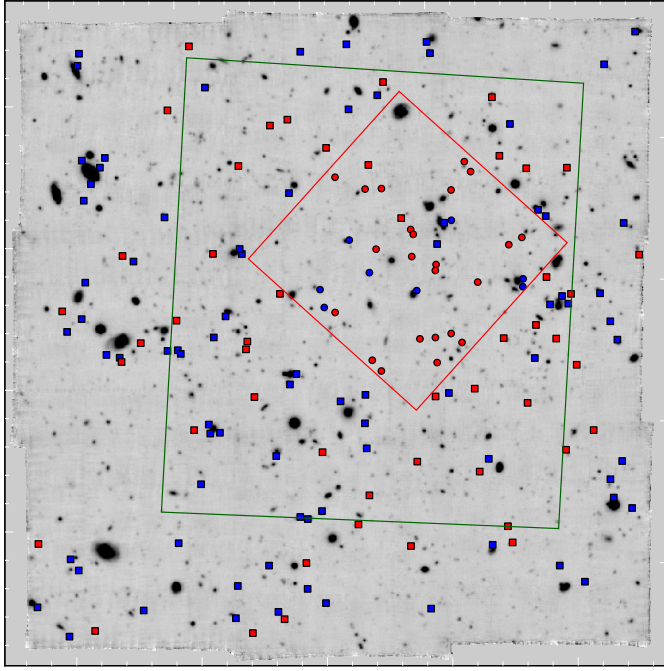
**Fig. 23.** On top, from left to right: ORIGIN MAXMAP image, MUSE reconstructed white light image, HST images in the *F775W* and *F850LP* filters, and  $\text{Ly}\alpha$  narrow-band image. Image size is  $5''$  and the source center is indicated by a red crosshair. The blue circles mark the sources identified in the Rafelski catalog. Bottom: source spectrum over the whole wavelength range (box-filtered with a window of 5 pixels) and zoomed (unfiltered) around the  $\text{Ly}\alpha$  line. The noise standard deviation is shown in magenta (mirrored with respect to the source spectra).

image is a flattened image of the correlation datacube, displaying the maximum of the correlation over wavelengths. The typical asymmetric  $\text{Ly}\alpha$  line profile is very clear, leading to a redshift of 6.24 for this object. Although the source was not identified in the Rafelski et al. (2015) catalog, a faint counterpart is present in the HST *F850LP* broadband image. The corresponding measured magnitude is  $\text{AB } 29.48 \pm 0.18$  (see Sect. 7.3).

The second object (Fig. 23b) is in the udf-10 field. It is also unambiguously detected by ORIGIN ( $P < 10^{-9}$ ). The line shape, although less asymmetric than for the previous case, the absence of other emission lines and the undetected continuum, qualifies the galaxy as a  $\text{Ly}\alpha$  emitter at  $z = 5.91$ , but this time one cannot see any HST counterpart. The derived lower limit magnitude is AB 30.7 in the corresponding *F850LP* broadband filter. In total ORIGIN detected 160 sources which were missed in the Rafelski

catalog, including 72 which have no HST counterpart (see next section).

We investigate how reliable is the detection of these 72 new sources by comparing their  $P$ -values with the corresponding values of the ORIGIN detections (restricted to  $\text{Ly}\alpha$  emitters) and successfully matched with an HST source. The histograms of the  $P$ -values for the two populations are given in Fig. 22. As expected, the sources with low  $P$ -values ( $< 10^{-29}$ ,  $< 10^{-18}$  in udf-10 and mosaic respectively) are all detected in HST. However, except for these bright emitters, the  $P$ -values of the HST-undetected sources are not very different from the general population. This is especially true for udf-10 which goes deeper than mosaic. At similar  $P$ -values, the sources detected by ORIGIN with HST counterpart were unambiguously identified (see Paper II for the detailed evaluation) giving confidence



**Fig. 24.** Location of the new sources detected by ORIGIN overlaid on the mosaic white-light image. *HST*-detected objects, i.e., brighter than the detection depth in at least one HST filter, are shown in blue, while the *HST*-undetected ones are displayed in red. The udf-10 and mosaic sources are marked with a circle and a square symbol, respectively. The green rectangle indicates the XDF/HUDF09/HUDF12 region containing the deepest near-IR observations from the HST WFC3/IR camera. The red square show the udf-10 field location. The north is located 42° clockwise from the vertical axis.

that most of the *HST*-undetected sources found by ORIGIN are real.

### 7.3. *HST* photometry of newly detected sources

We performed a simple aperture photometric analysis by computing HST AB magnitudes in a 0".4 diameter centered at the source location for all HST broadband images<sup>13</sup>. The magnitudes were compared to the 5 $\sigma$  detection limit of the corresponding HST filter (see column AB<sub>5 $\sigma$</sub>  in Table 2). A source is defined as *HST*-detected when it is brighter than the 5 $\sigma$  detection limit in at least one of the HST filters. Note that for the sources which fall outside the region with the deepest WFC3 IR data the corresponding shallower limiting depth was used.

The location of all sources without Rafelski et al. (2015) catalog entries are shown in Fig. 24. Among these 160 sources, 72 were considered as *HST*-undetected, i.e., with all computed magnitude larger than the detection limit. While the majority of these objects (54) are located within the region with the deepest WFC3 IR data, a small fraction (18) are found outside this region. Although all of these objects without HST counterpart fall below the detection limit of Rafelski, we derive a rough estimate of their average magnitude by computing the mean AB magnitude and its standard deviation for the entire sample of 54 sources present in the area of the HST deepest IR images (Table 2). A detailed analysis of the properties of these sources is deferred to another paper (Maseda et al., in prep.).

<sup>13</sup> Note that these fixed aperture magnitudes can be different from those given in Paper II which are based on the NoiseChisel segmentation maps.

**Table 2.** Mean HST AB magnitude ( $\overline{AB}$ ) of the 54 sources without HST counterpart in the deepest UDF region (displayed as a green rectangle in Fig. 24).

Filter	$\overline{AB}$	AB <sub>5<math>\sigma</math></sub>
<i>F606W</i>	31.8 $\pm$ 1.1	29.6
<i>F775W</i>	31.3 $\pm$ 1.3	29.5
<i>F850LP</i>	31.0 $\pm$ 1.4	28.9
<i>F105W</i>	31.3 $\pm$ 0.9	30.1
<i>F125W</i>	31.1 $\pm$ 0.6	29.7

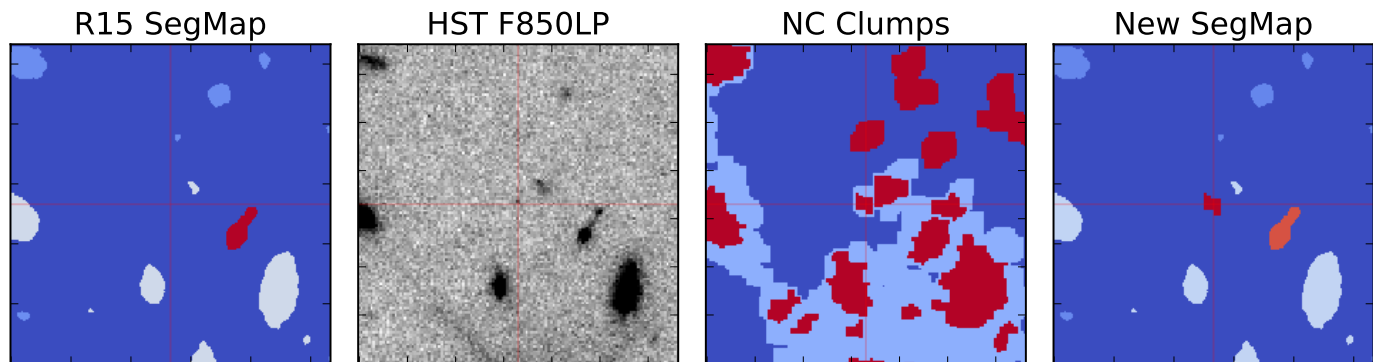
**Notes.** The reference AB 5 $\sigma$  depth (AB<sub>5 $\sigma$</sub> ) from Table 1 of Rafelski et al. (2015) is shown.

We inspect the 88 *HST*-detected objects discussed above to understand why they were missing in that catalog. We found three main reasons: 1) distant deblending, where the object is clearly detected but parametric fitting had associated it with a distant neighbor, see Fig. 17 in Akhlaghi & Ichikawa (2015); 2) nearby deblending, where the object was too close to a bright object to be identified as a separate object; and 3) manual removal based on S/N after running SExtractor, to correct for low purity. These three classes constituted 8%, 73%, and 15% of the missed objects.

To perform optimal source extraction as presented in Sect. 7.1, we update the Rafelski segmentation map with the segments corresponding to the new detected object. Rafelski et al. (2015) had already used multiple SExtractor (Bertin & Arnouts 1996) runs to generate their segmentation map. Hence for image segmentation and broadband measurements of these objects, we adopted NoiseChisel (Akhlaghi & Ichikawa 2015). NoiseChisel is non-parametric and much less sensitive to the diffuse flux of the neighboring objects. Therefore it is ideally suited to complement the Rafelski et al. (2015) catalog.

NoiseChisel was configured to “grow” the detected “clumps” into the diffuse regions surrounding them when there are no other clumps (resolved structure) over the detection area (see Fig. 10 of Akhlaghi & Ichikawa 2015). The final segmentation map for each object was selected as the one which gives the largest detection area among all filters. Checking the correspondence between magnitudes derived with this configuration and with Rafelski et al. (2015), we found the expected agreement in derived magnitudes: that is, in the AB magnitude interval 27.5 $\pm$ 0.25, the 2 $\sigma$  iterative clipped rms (terminated when the relative change in rms goes below 0.1) was 0.13 in the *F775W* filter. As a comparison, the R15 catalog has rms of 0.14 with the same magnitude interval, filter and method.

NoiseChisel detected the previously mentioned objects, along with another 39% of the initial sample. For the remaining objects, an aperture of diameter 0".5 was placed on the position reported by ORIGIN. Each object’s footprint was randomly placed in 200 non-detected regions and the 1 $\sigma$  width of the final distribution was defined as an upper limit on the magnitude. In the case of the WFC3/IR images that contain the wide HUDF and deep XDF/IR depths, this was done on the depth the object was positioned in, not the full UDF area. When the object’s magnitude was below the upper limit magnitude in a filter, the latter was used in the catalog. An example of a NoiseChisel detection performed on one of the sources without a Rafelski et al. (2015) catalog entry (ID 6524 Fig. 23a) is presented in Fig. 25.



**Fig. 25.** Complementing the Rafelski et al. (2015) segmentation map with NoiseChisel on source ID 6524 (see also Fig. 23a). Note that images are displayed in the original HST grid (rotated by  $-42^\circ$  compared to Fig. 23a). Image size is  $8''$  and the target is in the center (shown by the red crosshair). From left to right: the Rafelski et al. (2015) segmentation map, the input *F850LP* image, NoiseChisel clumps (red) over diffuse detections (light blue), and the final segmentation map, with the central clump of the previous image added to the input segmentation map. Note how some red regions in the NoiseChisel clumps image are not surrounded by diffuse flux (light blue). The measured magnitude is  $29.49 \pm 0.18$  in the *F850LP* filter. See Sect. 7.3 for more details.

## 8. Summary and conclusion

In this first paper of the series, we have presented the MUSE observational campaign of the *Hubble* Ultra Deep Field for a total of 137 h of VLT time, performed in 2014 and 2015 over eight runs of our GTO. A contiguous area of  $9.92 \text{ arcmin}^2$  was observed with a mosaic of nine fields. It covers almost the entire UDF region at a median depth of 9.6 h. A single field (udf-10) of  $1.15 \text{ arcmin}^2$  located within the XDF region, was also observed at additional depth. When combined with the mosaic fields, it reaches a median depth of 30.8 h.

The reduction of this large data set was performed using an advanced scheme to better remove the systematics and improve the overall quality of the produced datacubes. An enhanced self-calibration process, a better masking of instrument artefacts and the use of the PCA ZAP (Soto et al. 2016) software to remove sky residuals, results in datacubes with improved quality with respect to the previous HDFS MUSE observations and data reduction (Bacon et al. 2015).

We investigated the astrometry and broadband photometric properties of the datacubes, using the HST deep images as reference. We found an astrometric accuracy of  $0''.07$  rms, i.e.,  $\frac{1}{10}$  of the spatial resolution, for galaxies brighter than AB 27. We also assessed the broadband photometric performance, still using HST magnitude as reference. Although the achieved photometric accuracy of MUSE datacubes cannot compete with the performance of the UDF HST deep broadband imaging, especially in the redder part of the spectrum dominated by OH lines, we found good agreement with little systematic offset up to magnitude AB 28. The scatter of MUSE magnitudes with respect to HST is 0.4 mag in *F606W* for the udf-10 datacubes at AB 26.5, and 0.8 mag for the *F775W* and *F814W* filters at the same magnitude.

We developed an original method to accurately measure the spatial resolution of the observations through a comparison with the HST broadband images. This method can be used when there is no bright star in the MUSE field. It works in Fourier space and also provides a good estimate of the absolute astrometric and photometric offsets with respect to HST. Using this new tool, we derived the spatial PSF of the combined datacubes, modeled as a Moffat function with a constant  $\beta = 2.8$  parameter and a linear decrease of FWHM with wavelength. The achieved spatial resolution (Fig. 14) is  $0''.71$  (at  $4750 \text{ \AA}$ ) and  $0''.57$  (at  $9350 \text{ \AA}$ )

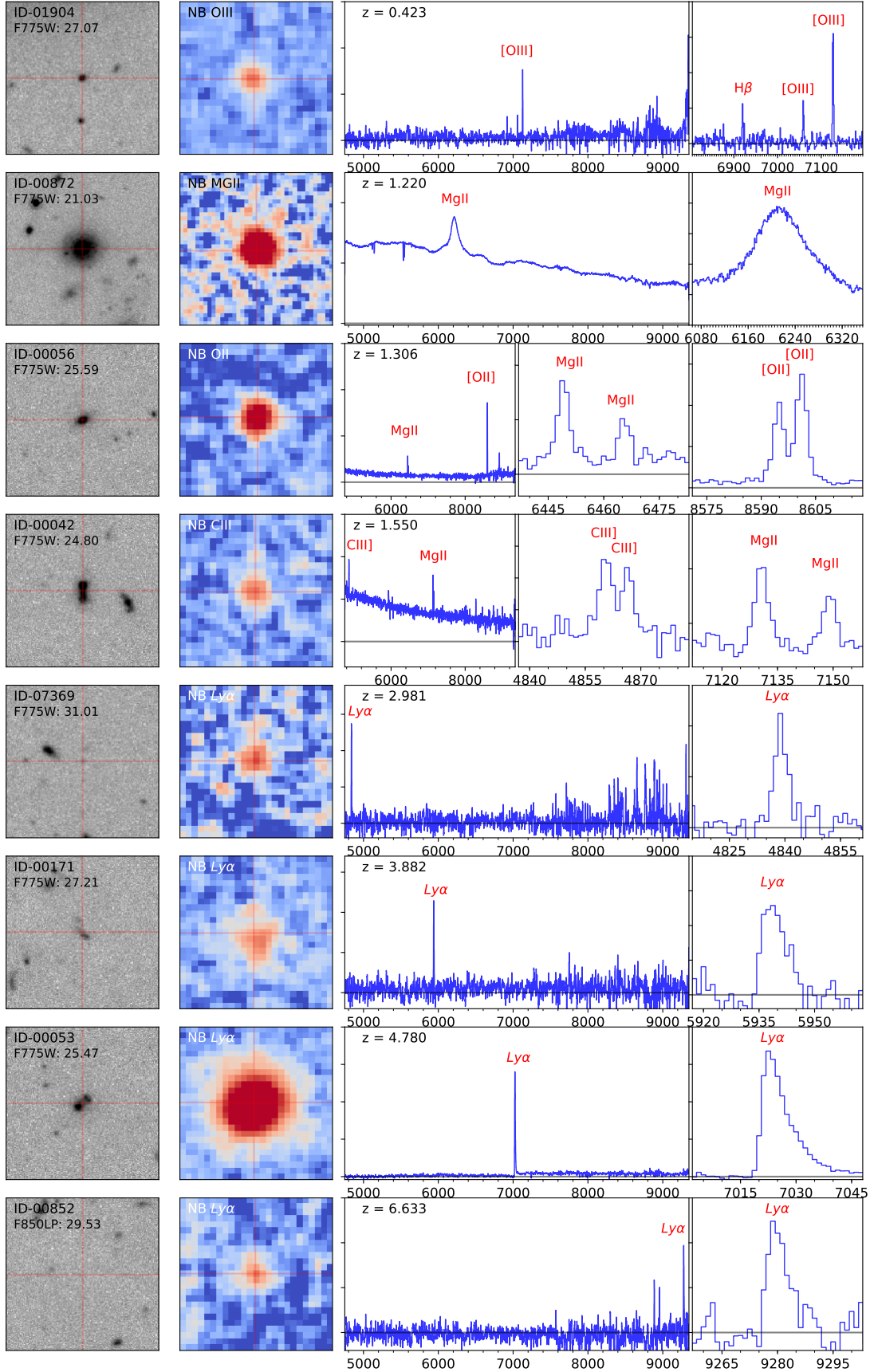
FWHM for both the mosaic and udf-10 fields. There is little dispersion for the mosaic sub-fields, with a measured standard deviation of only  $0''.02$ .

We investigated the noise properties of the two final datacubes. The noise distribution is well represented by a Normal probability density function. The empirical correction accounting for the correlated noise in each individual datacube prior to the combination works well. The final corrected propagated standard deviation is a good representation of the true noise distribution in regions with faint sources (e.g., dominated by the sky noise). A  $1\sigma$  surface brightness emission line sensitivity (Fig. 19) of  $2.8$  and  $5.5 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  is reached in the red for an aperture of  $1'' \times 1''$  and for the udf-10 and mosaic datacubes, respectively. This is a factor 1.6 better than the sensitivity measured in the first release of the HDFS datacube, demonstrating the progress achieved in the data reduction and observational strategy. A  $3\sigma$  point source line detection limit (Fig. 20) of  $1.5$  and  $3.1 \times 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2}$  is achieved in the red ( $6500\text{--}8500 \text{ \AA}$ ) and between OH sky lines for the udf-10 and mosaic datacubes, respectively.

We extracted 6288 and 854 sources from the mosaic and udf-10 datacubes, using the Rafelski et al. (2015) catalog and segmentation map as input for the source locations. For each source we performed optimal extraction, weighted with either the white light image or the FSF at the source location. A large number (40%) of HST sources are blended at the MUSE spatial resolution, but we show that this blending can often be resolved using reconstructed narrow-band images to locate sources that have detected emission lines.

In parallel we performed a blind search for emission line objects using an algorithm (ORIGIN) developed specially for MUSE datacubes. ORIGIN computes test statistics on a matched filtered datacube after a PCA-based continuum removal. The blind search results in 306 and 1251 detections in the udf-10 and mosaic datacubes, respectively.

A number of these sources (160) were not present in the Rafelski et al. (2015) catalog. Investigation of these new sources show that 55% of them are bright enough in at least one of the HST band to be detected, but have been missed because of contamination and/or uncorrect SExtractor deblending process. The remaining 72 sources fall below the detection limit of HST broadband deep images. In the HST region with deep WFC3/IR images, we compute a mean AB magnitude of  $31.0\text{--}31.8$  within



**Fig. 26.** Example of sources from the mosaic and udf-10 fields. Each row shows a different object, ordered by redshift. *From left to right* one can see: the HST broadband image (*F775W* filter), a MUSE-reconstructed narrow-band image of one of the brightest emission lines, the source spectrum over the full wavelength range and a zoom-in region highlighting some characteristic emission lines. The images have a linear size of 5'' and the source center is displayed as a red cross-hair.

a 0.4 diameter aperture. We use NoiseChisel, a SExtractor alternative optimized for the detection of diffuse sources, to derive an updated segmentation map for these sources when that was possible.

The redshift measurement and analysis of this unprecedented data set is presented in Paper II (Inami et al. 2017). With more than 1300 high-quality redshifts, this survey is the deepest and most comprehensive spectroscopic study of the UDF ever performed. It expands the present spectroscopy data set (173 galaxies accumulated over ten years) by almost an order of magnitude and covers a wide range of galaxies, from nearby objects to  $z = 6.6$  high redshift Ly $\alpha$  emitters, and from bright (magnitude 21) galaxies to the faintest objects (magnitude >30) visible in the HST images.

Of course, the survey “performance” is much more than just the number of faint sources from which we are able to obtain reliable redshifts. The quality of the MUSE data, as shown in Fig. 26 for a few representative sources, enables new and detailed studies of the physical properties of the galaxy population and their environments over a large redshift range. In subsequent papers of this series, we will therefore explore the science content of this unique data set.

*Acknowledgements.* R.B., S.C., H.I., J.B.C., M.S., M.A. acknowledge support from the ERC advanced grant 339659-MUSICOS. J.R., D.L. acknowledge support from the ERC starting grant CALENDs. R.B., T.C., B.G., N.B., B.E., J.R. acknowledge support from the FOGHAR Project with ANR Grant ANR-13-BS05-0010. J.S. acknowledge support from the ERC grant 278594-GasAroundGalaxies. S.C. and R.A.M. acknowledge support from Swiss National Science Foundation grant PP00P2 163824. Part of this work was granted access to the HPC and visualization resources of the Centre de Calcul Interactif hosted by University Nice Sophia Antipolis. B.E. acknowledges financial support from “Programme National de Cosmologie et Galaxies” (PNCG) of CNRS/INSU, France. J.B. acknowledges support by Fundação para a Ciência e a Tecnologia (FCT) through national funds (UID/FIS/0434/2013) and Investigador FCT contract IF/01654/2014/CP1215/CT0003, and by FEDER through COMPETE2020 (POCI-01-0145-FEDER-007672). P.M.W. received support through BMBF Verbundforschung (project MUSE-AO, grant 05A14BAC).

## References

- Akhlaghi, M., & Ichikawa, T. 2015, *ApJS*, **220**, 1
- Aravena, M., Decarli, R., Walter, F., et al. 2016a, *ApJ*, **833**, 71
- Aravena, M., Decarli, R., Walter, F., et al. 2016b, *ApJ*, **833**, 68
- Bacon, R., Bower, R., Cabrit, S., et al. 2004, MUSE Science Case (ESO Report)
- Bacon, R., Accardo, M., Adjali, L., et al. 2010, in *SPIE Conf. Ser.*, **7735**, 8
- Bacon, R., Vernet, J., Borosiva, E., et al. 2014, *The Messenger*, **157**, 21
- Bacon, R., Brinchmann, J., Richard, J., et al. 2015, *A&A*, **575**, A75
- Balestra, I., Mainieri, V., Popesso, P., et al. 2010, *A&A*, **512**, A12
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, **132**, 1729
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Borisova, E., Cantalupo, S., Lilly, S. J., et al. 2016, *ApJ*, **831**, 39
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, *ApJ*, **737**, 90
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, **803**, 34
- Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, *ApJ*, **833**, 72
- Brinchmann, J., Inami, H., Bacon, R., et al. 2017, *A&A*, **608**, A3 (MUSE UDF SI, Paper III)
- Carton, D., Brinchmann, J., Shirazi, M., et al. 2017, *MNRAS*, **468**, 2140
- Comastri, A., Ranalli, P., Iwasawa, K., et al. 2011, *A&A*, **526**, L9
- Conseil, S., Bacon, R., Piqueras, L., & Shepherd, M. 2017, in ADASS XXVI (held October 16–20, 2016) Proc., in press [[arXiv:1612.05308](https://arxiv.org/abs/1612.05308)]
- Conselice, C. J., Bluck, A. F. L., Ravindranath, S., et al. 2011, *MNRAS*, **417**, 2770
- Contini, T., Epinat, B., Bouché, N., et al. 2016, *A&A*, **591**, A49
- Curtis-Lake, E., McLure, R. J., Dunlop, J. S., et al. 2016, *MNRAS*, **457**, 440
- Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, *ApJ*, **626**, 680
- Decarli, R., Walter, F., Aravena, M., et al. 2016a, *ApJ*, **833**, 70
- Decarli, R., Walter, F., Aravena, M., et al. 2016b, *ApJ*, **833**, 69
- Drake, A. B., Guiderdoni, B., Blaizot, J., et al. 2017a, *MNRAS*, **471**, 267
- Drake, A., Garel, T., Hashimoto, T., et al. 2017b, *A&A*, **608**, A6 (MUSE UDF SI, Paper VI)
- Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, *MNRAS*, **466**, 861
- Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, *ApJ*, **763**, L7
- Epinat, B., Tasca, L., Amram, P., et al. 2012, *A&A*, **539**, A92
- Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., et al. 2015, *ApJ*, **810**, 71
- Finley, H., Bouché, N., Contini, T., et al. 2017a, *A&A*, **605**, A118
- Finley, H., Bouché, N., Contini, T., et al. 2017b, *A&A*, **608**, A7 (MUSE UDF SI, Paper VII)
- Fumagalli, M., Cantalupo, S., Dekel, A., et al. 2016, *MNRAS*, **462**, 1978
- González, V., Labbé, I., Bouwens, R. J., et al. 2011, *ApJ*, **735**, L34
- Grazian, A., Fontana, A., Santini, P., et al. 2015, *A&A*, **575**, A96
- Guérou, A., Krajnovic, D., Epinat, B., et al. 2017, *A&A*, **608**, A5 (MUSE UDF SI, Paper V)
- Hashimoto, T., Garel, T., Guiderdoni, B., et al. 2017, *A&A*, **608**, A10 (MUSE UDF SI, Paper X)
- Herenz, E. C., & Wisotzki, L. 2017, *A&A*, **602**, A111
- Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, *A&A*, **606**, A12
- Horne, K. 1986, *PASP*, **98**, 609
- Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, *ApJS*, **209**, 6
- Inami, H., Bacon, R., Brinchmann, J., et al. 2017, *A&A*, **608**, A2 (MUSE UDF SI, Paper II)
- Kay, S. M. 1998, *Fundamentals of Statistical Signal Processing: Detection Theory*, Vol. 2 (Prentice-Hall PTR)
- Kellermann, K. I., Fomalont, E. B., Mainieri, V., et al. 2008, *ApJS*, **179**, 71
- Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, *ApJS*, **209**, 3
- Kurk, J., Cimatti, A., Daddi, E., et al. 2013, *A&A*, **549**, A63
- Labbé, I., Oesch, P. A., Illingworth, G. D., et al. 2015, *ApJS*, **221**, 23
- Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, *A&A*, **608**, A8 (MUSE UDF SI, Paper VIII)
- Le Fèvre, O., Vettolani, G., Paltani, S., et al. 2004, *A&A*, **428**, 1043
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, *ApJS*, **228**, 2
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Maseda, M., Brinchmann, J., Franx, M., et al. 2017, *A&A*, **608**, A4 (MUSE UDF SI, Paper IV)
- McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, *MNRAS*, **432**, 2696
- Meillier, C., Chatelain, F., Michel, O., et al. 2016, *A&A*, **588**, A140
- Mignoli, M., Cimatti, A., Zamorani, G., et al. 2005, *A&A*, **437**, 883
- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, *ApJS*, **225**, 27
- Morris, A. M., Kocevski, D. D., Trump, J. R., et al. 2015, *AJ*, **149**, 178
- Oesch, P. A., Bouwens, R. J., Carollo, C. M., et al. 2010, *ApJ*, **709**, L21
- Ono, Y., Ouchi, M., Curtis-Lake, E., et al. 2013, *ApJ*, **777**, 155
- Paris, S., Suleiman, R., Mary, D., & Ferrari, A. 2013, in Proc. ICASSP 2013
- Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, *MNRAS*, **456**, 3194
- Pearson, E. S., D’Agostino, R. B., & Bowman, K. O. 1977, *Biometrika*, **64**, 231
- Piqueras, L., Conseil, S., Shepherd, M., et al. 2017, in ADASS XXVI (held October 16–20, 2016) Proc., in press [[arXiv:1710.03554](https://arxiv.org/abs/1710.03554)]
- Popesso, P., Dickinson, M., Nonino, M., et al. 2009, *A&A*, **494**, 443
- Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, *AJ*, **150**, 31
- Rafelski, M., Gardner, J. P., Fumagalli, M., et al. 2016, *ApJ*, **825**, 87
- Rujopakarn, W., Dunlop, J. S., Rieke, G. H., et al. 2016, *ApJ*, **833**, 12
- Scharf, L. L., & Friedlander, B. 1994, *IEEE Trans. on Signal Processing*, **42**, 2146
- Shibuya, T., Ouchi, M., & Harikane, Y. 2015, *ApJS*, **219**, 15
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, *ApJ*, **825**, 5
- Soto, K. T., Lilly, S. J., Bacon, R., Richard, J., & Conseil, S. 2016, *MNRAS*, **458**, 3210
- Szokoly, G. P., Bergeron, J., Hasinger, G., et al. 2004, *ApJS*, **155**, 271
- Szomoru, D., Franx, M., Bouwens, R. J., et al. 2011, *ApJ*, **735**, L22
- Teplitz, H. I., Rafelski, M., Kurczynski, P., et al. 2013, *ApJ*, **146**, 159
- Tokovinin, A. 2002, *PASP*, **114**, 1156
- van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014, *ApJ*, **792**, L6
- Vanzella, E., Cristiani, S., Dickinson, M., et al. 2005, *A&A*, **434**, 53
- Vanzella, E., Cristiani, S., Dickinson, M., et al. 2006, *A&A*, **454**, 423
- Vanzella, E., Cristiani, S., Dickinson, M., et al. 2008, *A&A*, **478**, 83
- Vanzella, E., Gialalisco, M., Dickinson, M., et al. 2009, *ApJ*, **695**, 1163
- Ventou, A., Contini, T., Bouché, N., et al. 2017, *A&A*, **608**, A9 (MUSE UDF SI, Paper IX)
- Voyer, E. N., de Mello, D. F., Siana, B., et al. 2009, *AJ*, **138**, 598
- Walter, F., Decarli, R., Aravena, M., et al. 2016, *ApJ*, **833**, 67
- Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, *A&A*, **587**, A98
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, **195**, 10